

**Time-Series Modelling in Financial Markets: New Approaches and  
Exchange Rate Applications**

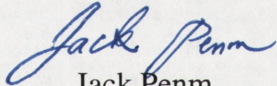
**by**

**Jack H.W. Penm**

**A thesis submitted for the degree of  
Doctor of Philosophy (Finance)  
of the Australian National University**

**December 2001**

In compliance with the requirements relating to Admission to Examination for the Degree of Doctor of Philosophy of the Australian National University, it is affirmed that, unless otherwise stated, the work that follows is my own.

  
Jack Penm

## **ACKNOWLEDGEMENTS**

I am grateful to my supervisors, Professor Tim Brailsford (Chair), Professor Deane Terrell and Professor Peter Hall for their support and supervision on this PhD thesis. I also wish to thank the School of Finance and Applied Statistics, especially the Head – Professor Terry O'Neill for the assistance I received during the course of this study. The School has been a stimulating environment for undertaking a course of this nature.

## TABLE OF CONTENTS

Acknowledgements	3
Table of Contents	4
List of Tables	6
List of Figures	8
Abstract	10
Chapter 1: Purpose and Scope of Study	12
Chapter 2: Price Movements in Foreign Exchange Markets	26
Chapter 3: New Modelling Techniques in Financial Time-series Analysis	52
Chapter 4: Causality Detection and Cointegration Investigation	85
Chapter 5: Causal Analysis of the Money Supply and the Euro and Its Impact on the Hong Kong Stock Market	103
Chapter 6: Purchasing Power Parity Tests in Foreign Exchange Markets	123
Chapter 7: Non-linear Forecasting Using Neural Networks	157
Chapter 8: Selecting the Forgetting Factor in Subset Autoregressive Modelling	190
Chapter 9: Summary and Conclusions	223



List of Appendices	229
Appendix A: Supplementary Mimeo to the Thesis	230
Appendix B: Tree-Pruning Algorithm for Cointegration	252
Appendix C: Identified Optimal VARs for Simulations	262
Appendix D: Understanding the Forgetting Factor via Kernel Regression	265
Bibliography	269

## LIST OF TABLES

### Chapter 5

5.1	Orthogonal polynomial regression	114
5.2	The VARs selected by both HQC and SC for detecting the causal relationships between $E_e$ and M3	116
5.3	The optimal ZNZ patterned VAR selected by SC	119
5.4	Causal pattern detected in the three variable system selected by SC	119
5.5	Causal relationships in the three variable system selected by SC	120

### Chapter 6

6.1	Fixed and flexible exchange rate regimes, and the Exchange Rate Mechanism	127
6.2	Test results for purchasing power parity of bilateral exchange rates from 1975(1) to 1994(4) using unit root tests	130
6.3	Estimated VECMs which support the necessary condition for PPP of OECD exchange	134
6.4	Estimated VECMs which support the necessary condition for PPP of Asian exchange rates	136
6.5	Test of the necessary and sufficient condition for PPP	139
6.6	The VECM for the relationship linking exchange rates and consumer price indices between Australia and the USA	149
6.7	The VECM for the relationship linking money supply, inflation and stock market indicator for Australia	154

## Chapter 7

7.1	The necessary VRDLs and the forward and backward VARs	177
7.2	The VRDLs selected by MHQC for detecting the causal relationship from SPI to AOI	183
7.3	The VRDLs selected by MHQC for detecting the causal relationship from AOI to SPI	184
7.4	ZNZ VAR, full-order VAR, and random-walk forecasts: Telstra share price	186

## Chapter 8

8.1	Order of the subset AR models determined using MHQC	214
8.2	Improvements over <i>ex ante</i> forecasts using the forgetting factor	215
8.3	The TRML estimation and the bootstrap results for the AOI	220

## Appendix C

C.1	The optimal subset VAR models selected for MHQC	262
-----	-------------------------------------------------	-----

## LIST OF FIGURES

### Chapter 2

2.1	Quarterly real effective exchange rate indices of the US Dollar (1980-2000)	41
2.2	Weekly nominal exchange rates of the Euro relative to the US Dollar (1999-2000)	43
2.3	Quarterly nominal exchange rates of the Japanese Yen relative to the US Dollar (1980-2000)	44
2.4	Quarterly trade-weighted exchange rate indices of the Australian Dollar (1980-2000)	47

### Chapter 3

3.1	An extended two-layered neural network	78
3.2	A polynomial neural network with a single hidden-layer for a VRDL model	81
3.3	A three-layered polynomial neural network for a VAR model	82

### Chapter 5

5.1	Growth of M3 in the Euro area	106
5.2	Euro to US Dollar - exchange rate (EUFX), daily: 1 January 1999 to 31 December 1999	112
5.3	Hang Seng stock price index (HSI), daily: 1 January 1999 to 31 December 1999	113
5.4	Hong Kong Dollar to US Dollar - exchange rate (HKFX), daily: 1 January 1999 to 31 December 1999	113



## Chapter 7

7.1	Log AOI in first differences, daily: 27 January 1998 to 29 June 1998	181
7.2	Log SPI in first differences, daily: 27 January 1998 to 29 June 1998	181
7.3	Log (Telstra's share price), daily: 24 November 1997 to 30 March 1998	186

## Chapter 8

8.1	The reciprocal integer pairs of a subset AR model with $K=4$	202
8.2	Recursions to acquire $\underline{g}$ and $\beta$ for each reciprocal integer pair of the forward and backward AR models ( $K=4$ )	210
8.3	Real exchange rates for the USA with Germany, United Kingdom and Korea, monthly: January 1975 to December 1993	213
8.4	The All Ordinaries Index levels series over the period 15 June 1995 to 14 May 1996	217

## Appendix B

B.1	A four-variable inverse tree	254
-----	------------------------------	-----

## **ABSTRACT**

This thesis investigates and develops time-series models in financial markets. It provides new approaches to this area, and applies these approaches to various exchange rate contexts. The aims of this thesis are: to develop an understanding of modelling international financial market movements, in particular the foreign exchange market and the Australian equity market; to examine advances in non-linear time-series modelling using computer-intensive statistical approaches; and to investigate and develop non-linear time-series modelling in financial markets.

The topic of time-series modelling in financial markets has received widespread coverage in the literature. The issue is important because of its fundamental role in investment decision-making. Over the last decade, financial markets have been affected by significant structural changes, including the expansion of financial markets, the globalisation of finance, the introduction of the Euro, the increasing role of electronic broking, and the massive growth in the investment funds industry. These factors have contributed to an increased focus on financial market movements. Conventional linear time-series models are unlikely to properly explain and model significant price variations. Non-linear time-series models have the potential to improve the performance of modelling and are appealing in explaining complex price behavior.

This thesis investigates time-series modelling in financial markets by introducing an innovative approach to model building. A range of new techniques for vector time-series modelling is used to enhance existing linear and non-linear modelling of exchange rate and equity market movements. These techniques include zero-non-zero (ZNZ) patterned vector autoregressive (VAR) modelling, ZNZ patterned vector error-correction (VECM) modelling, ZNZ patterned polynomial neural networks, the forgetting factor method and bootstrapping. Traditional full-order models assume all non-zero entries in their coefficient matrices. This thesis develops a vector time-series model, with allowance for possible zero entries in coefficient matrices, as a ZNZ patterned vector time-series model.

The model's optimal order and ZNZ pattern determination can be drawn from procedures using artificial intelligence techniques such as structured search algorithms. If key linear and non-linear interactions among variables are captured accurately, the chosen ZNZ system can improve the modelling and simulation performance. While no model will ever completely capture the characteristics of complex modern financial systems, the combination of these approaches provides important insights into, and improvements in the performance of, statistical procedures in financial time-series analysis.

# **CHAPTER 1**

## **PURPOSE AND SCOPE OF STUDY**

### **1.1 Aims**

This thesis examines the underlying time-series forces that contribute significantly to price movements in international forex and financial markets. The aims of this thesis are: to develop an understanding of modelling price movements, in particular the foreign exchange market and the Australian equity market; to examine advances in non-linear time-series modelling using computer-intensive statistical approaches; and to investigate and develop the performance of non-linear time-series modelling in financial markets.

Model building is a major focus of this thesis. New technology for non-linear modelling is helpful in interpreting and explaining certain aspects of the price behaviour of financial markets.

### **1.2 Linear and Non-linear Modelling**

While the traditional linear modelling approach used to investigate price movements has received widespread coverage in the literature, this thesis takes a new approach to modelling. The use of non-linear models in statistical and econometric research has increased in recent years [see Refenes *et al* (1995)].



Non-linear models can improve the performance of modelling due to the flexibility of such models in accounting for potentially complex non-linear relationships not captured by linear models. At the same time they encompass simulation instruments which can be used for analysing market properties. For instance, they are important in explaining and modelling cycles characterised by flat bottoms and sharp peaks. These characteristics are commonly observed in exchange rate and equity price markets, especially when the relevant data are recorded at relatively high frequencies. These non-linear models can also subsume in their specifications, linear relationships supported by theory, such as long-term linear cointegrating relations.

Over the last decade financial markets have experienced many important changes. The expansion of financial markets, the globalisation of capital markets, the massive growth in the investment funds industry, the introduction of the Euro and the integration of foreign exchange markets are all factors that have contributed to an increased focus on financial market price movements. Adrangi *et al* (2001) claim that linear models are unlikely to properly capture sudden movements and wide fluctuations in these markets. The existence of non-linearities in financial market movements has been emphasized by various researchers. Frank and Stengos (1989) find evidence of non-linear structures in the rates of return of gold and silver. Brooks and Henry (2000) demonstrate linear and non-linear transmission of equity return volatility among the United States, Japanese and Australian markets. De Grauwe *et al* (1993) show there exists a non-linear mechanism that drives exchange rate series and provide evidence of these non-linearities.

This thesis develops a vector time-series model, with allowance for possible zero entries in coefficient matrices, as a zero-non-zero (ZNZ) patterned vector time-series model. Commonly employed full-order vector time-series models assume nonzero entries in all their coefficient matrices. A range of new techniques for vector time-series modelling is used in the thesis to enhance existing non-linear modelling of exchange rate and equity markets. These techniques include ZNZ patterned vector autoregressive (VAR) modelling, ZNZ patterned vector error-correction (VECM) modelling, ZNZ patterned polynomial neural networks, the forgetting factor method and bootstrapping. The model's optimal order and ZNZ pattern determination can be drawn from procedures using artificial intelligence techniques such as structured search algorithms. These new techniques bring new insights to understanding price behaviour in financial markets.

### **1.3 Major Areas of Interest**

In outline, this thesis investigates both modelling issues and applications. Modelling issues comprise ZNZ patterned VAR modelling, ZNZ patterned VECM modelling, neural networks and the forgetting factor method. Two applications are utilised to demonstrate the power of neural networks, and two applications are used to show the effectiveness of the forgetting factor method. Major applications include the relationship between foreign exchange and stock markets, and purchasing power parity (PPP) tests.

### **1.3.1 Modelling Issues**

#### **1.3.1.1 ZNZ Patterned VAR and VECM Modelling**

VAR models are increasingly being used in the analysis of relationships within and between financial markets. In such models there are circumstances that require zero entries in the coefficient matrices. Specifically, if it is necessary to detect the presence or absence of indirect causality and/or Granger non-causality in the framework of VAR, the efficiency of the causality detection is crucially dependent upon finding these zero coefficient entries where the true structure does indeed include zero entries. Such circumstances can be particularly relevant in the context of markets with special characteristics, such as emerging economies. This thesis shows that a direct extension of the use of the Yule-Walker relations for fitting VAR models with ZNZ patterned coefficient matrices is inconsistent with statistical procedures, as the resultant estimated variance-covariance matrix of the white noise disturbance process becomes non-symmetric. This inconsistency can lead to an inability to test financial theory. Chapter 5 provides a consistent adjustment which fits with the theory.

In Chapter 5 the practical use of the adjustment is demonstrated in a vector system comprising variables from the Hong Kong stock market and foreign exchange markets. The results indicate that the Euro leads the Hong Kong stock market. Since Hong Kong is an open market, capital flows to and from Europe make up a key component of trading in that stock market. The findings are helpful in explaining linkages between the financial variables involved.

In vector time-series analysis, VECMs have become an important means of detecting Granger causal relations and cointegrating relations. Commonly employed full-order VECM models assume nonzero entries in all their coefficient matrices. However, applications of VECM models to economic and financial time-series data have revealed that zero entries are indeed possible. The existence of zero entries has not been fully explored in causality and cointegration theory. Specifically, if indirect causality or Granger non-causality exists among the variables, the use of ‘overparameterised’ full-order VECM models may weaken the power of statistical inferences. This thesis argues that the ZNZ patterned VECM is a more straightforward and effective means of testing for both indirect causality and Granger non-causality. The same benefits will be present if the ZNZ patterned VECM is used to analyse cointegrating relations. Chapter 6 presents applications that demonstrate the usefulness of the ZNZ patterned VECM. The chapter indicates this new methodology is useful in providing insights into PPP.

### **1.3.1.2 Neural Networks**

Conventionally linear time-series approaches have been adopted in modelling financial time-series. As described in Section 1.2, the modelling power of linear approaches is weak in relation to the complexities of financial markets. This thesis focuses on non-linear models to improve performance in modelling and simulation. It is proposed that non-linear models, in particular neural networks, have the ability to improve the performance of modelling and simulating the



movements of financial variables, including equity market indicators and exchange rates. These networks have the flexibility to account for potentially complex non-linear relationships which cannot be fully captured by linear models.

Chapter 7 presents two applications of the newly developed learning algorithm for multi-layer neural networks. The first concerns a causality relationship between the stock and futures markets. The second concerns the relationship between prices of an individual share and the underlying stock market.

Further, this thesis extends the relevance of multi-layered neural networks and so more effectively models a greater array of financial vector time-series situations. It thus recognises that many connections between nodes in layers are unnecessary and can be deleted. This thesis introduces inhibitor arcs - reflecting inhibitive synapses. It also allows for connections between nodes in layers which have variable strengths at different points of time by introducing additionally excitatory arcs - reflecting excitatory synapses. In summary this innovative and sophisticated learning algorithm is simple to use and can avoid cumbersome matrix inversion, and therefore results in better numerical accuracy. The findings reveal that both the modelling and simulation performance can be improved by the chosen optimal specification.

### 1.3.1.3 The Forgetting Factor Method

In time-series modelling, the forgetting factor method assesses each incoming observation and applies appropriate weights to update the model structure and parameters. Although the use of the forgetting factor method in time-series modelling has grown, the procedure for determining its nature has not yet been fully investigated. This thesis provides further insight into how to characterise the forgetting factor in time-series analysis.

Conventional methods for determining the forgetting factor in autoregressive (AR) models are mostly based on arbitrary or personal choices. Chapter 8 presents two procedures which can be used to select the forgetting factor in subset AR modelling. The first procedure uses the bootstrap to determine the value of a fixed forgetting factor. The second procedure utilises the time recursive maximum likelihood (TRML) method, in conjunction with the bootstrap, to estimate the value of a dynamic forgetting factor. In two applications using exchange rate and stock market data, the effect on *ex ante* forecasting of both the fixed and the dynamic forgetting factors in subset AR modelling of non-stationary time-series is demonstrated.

### **1.3.2 Major Applications**

ZNZ patterned modelling is introduced in Chapter 3, and developed in Chapter 4 where it is applied to forex market movements. Two major applications are conducted. The first examines the relationship between foreign exchange and stock markets. The second investigates PPP testing.

#### **1.3.2.1 Relationship between Foreign Exchange and Stock Markets**

The interaction between foreign exchange markets and stock markets is generally driven by capital flows. Investors seek growth in stock markets and gains from movements in exchange rates. One application investigates the causal analysis between the money supply and the Euro. The Euro's impact on the Hong Kong stock market is also assessed. The ZNZ patterned VAR modelling is utilised to assess whether direct Granger causal relations exist between the money supply and the Euro's exchange rate with the US Dollar. Further, the ZNZ patterned VAR modelling is used to investigate the Euro's impact on the Hong Kong stock market during 1999.

Chapter 5 reports on the causal analysis which tests whether the Euro contributes to the movements in the Hong Kong stock market. The search algorithm proposed in Penm and Terrell (1984a), in conjunction with model selection criteria, is used to select the optimal ZNZ patterned VAR model. The selected optimal model is then used as a basis for Granger causal detection. The findings are consistent with economic theory and prior evidence. The Euro is detected as a variable which

produces leading information for the Hong Kong market. That is, a shock to the Euro causes a response in the Hong Kong market.

### **1.3.2.2 Purchasing Power Parity Testing**

This thesis also investigates the relevance of PPP hypothesis in explaining bilateral exchange rates between Australian and foreign currencies. PPP theory states that movements in the exchange rate between two countries' currencies are determined by movements in their relative prices.

PPP theory has important implications for exchange rate predictions, financial economic modelling and economic policy. If, for example, PPP is shown to be valid in the long-term, then the long-term relation between the nominal exchange rate and the ratio of domestic to foreign prices can be incorporated as a long-term restriction in a simulation model for the exchange rate. This inclusion would improve the accuracy of exchange rate predictions over the long-term, since any short-term deviation of the exchange rate from PPP can be modelled as an error-correction mechanism, requiring adjustments in the exchange rate for the long-term relationship to be restored.

In this thesis, PPP hypothesis is tested for a number of bilateral exchange rates, using ZNZ patterned VECM modelling. The ZNZ patterned VECM approach provides a unique functionality which can detect both Granger-causality and cointegrating relationship within the optimal VECM in an effective and straightforward manner. This approach is different from the conventional error-



correction modelling which could result in misleading inferences and inferior projections. The findings indicate that ZNZ patterned VECM modelling can accommodate both long-term and short-term responses. In addition, the current assessment of the short-term movements in the exchange rate within the framework of patterned VECM can influence the presence or absence of the long-term PPP relationship. An algorithm for an  $I(2)$  analysis is also developed to conduct PPP testing. The results support the presence of PPP.

#### **1.4 The Structure of the Thesis**

The structure of the thesis is as follows. In Chapter 2, price movements in foreign exchange markets are discussed. Chapter 3 contains an overview of the models in financial time-series analysis which will be used. Chapter 4 presents the use of ZNZ patterned VAR and VECM modelling to examine Granger causal relationships and cointegrating relationships. In Chapter 5, a causal analysis of a dominating factor influencing the Euro is conducted. This chapter discusses the implications of the Euro exchange rate for the Hong Kong forex market and the Hong Kong stock market using ZNZ patterned VAR modelling. In Chapter 6 the necessary condition, and the necessary and sufficient condition, for PPP are sequentially tested for the Australian and selected foreign exchange markets in a framework of patterned VECM for  $I(1)$  integrated time-series. An  $I(2)$  algorithm for patterned VECM modelling is also developed, and applied to PPP testing and stock market analysis.

Chapter 7 grapples with the difficult but important question of non-linear simulations, where multi-layered neural networks are used to increase non-linear modelling power for ZNZ patterned financial time-series systems. To demonstrate the usefulness of multi-layered neural networks, the developed modelling algorithm is applied to the relationship between the All Ordinaries Index and the Share Price Index Futures Contract, and is applied to another illustration examining the relationship between prices of an individual share and the underlying stock market. Chapter 8 introduces and develops the forgetting factor method in subset AR modelling. It proposes two procedures to determine the value of the forgetting factor. The first uses the bootstrap to select the value of a fixed forgetting factor. The second utilises both the bootstrap and the TRML estimation to estimate the value of a dynamic factor. The findings show that the forgetting factor however determined can improve the forecasting performance. Chapter 9 summarises the thesis and presents major conclusions.

### 1.4.1 Contributions

The major theoretical contributions and empirical findings of this thesis are as follows:

#### *Theoretical contributions:*

- (1) Conventional Yule-Walker relations for ZNZ patterned VAR modelling lead to a theoretical inconsistency. An adjustment to the Yule-Walker relations, which removes this theoretical inconsistency, is supplied.
- (2) Identical Granger-causality, Granger non-causality and indirect causality relations among the variables can be detected by ZNZ patterned VAR models or by equivalent VECM models.
- (3) Both I(1) and I(2) identification and estimation algorithms are developed to conduct cointegration analysis when dealing with a ZNZ patterned VECM. A VECM can accommodate both long-term and dynamic responses.
- (4) Incorporation of polynomial neural networks into ZNZ patterned modelling. Both innovative order-update and time-update identification and estimation algorithms are developed and explored.

(5) Incorporation of the bootstrap and the time-update recursive maximum likelihood (TRML) methods into subset AR modelling to decide the value of the forgetting factor. The linkage between kernel estimation and the forgetting factor method is established.

***Empirical findings:***

- (1) Money supply is identified as an important factor influencing the Euro.
- (2) The value of the Euro contributes significantly to the movements in the Hong Kong stock market. A shock to the Euro foreign exchange market impacts on the movements of both Hong Kong stock and foreign exchange (forex) markets.
- (3) The conventional single-equation unit roots tests for PPP do not achieve consistent results.
- (4) PPP is examined for fourteen exchange rates using I(1) analysis. Half of the rates support the necessary condition for PPP. Also three of the seven exchange rates investigated support the necessary and sufficient condition for PPP.
- (5) Tests of PPP and a three-variable system concerning the stock market use I(2) analysis. Support for the necessary condition for the PPP hypothesis linking the bilateral exchange rate between the Australian Dollar and US Dollar is confirmed.

- (6) The assessment of the short-term movements in the exchange rate within the framework of a patterned VECM influences the finding of the presence or absence of the long-term PPP relationship.
- (7) The inter-relationships among the stock market, money supply and inflation are detected using  $I(2)$  analysis. The results are consistent with both theory and prior evidence.
- (8) The causal relationships between AOI and SPI are detected using an extended two-layered neural network. The existence of instantaneous causal and bidirectional feedback relationships are detected.
- (9) The non-linearities in the relationship between share prices and the behaviour measure of the underlying stock market are examined, using a three-layered neural network. The resultant forecasting results outperform the conventional full-order VAR and naive random-walk forecasts.
- (10) Three real exchange rates series and the AOI series are used to demonstrate the effectiveness of the forgetting factor method in forecasting performance. The outcome indicates that the bootstrap is a reliable procedure for selecting the value of a fixed forgetting factor, and the TRML method, in conjunction with the bootstrap, is an effective procedure for estimating the value of a dynamic forgetting factor.

## **CHAPTER 2**

### **PRICE MOVEMENTS IN FOREIGN EXCHANGE MARKETS**

#### **2.1 Introduction**

The foreign exchange (forex) market is important because it is the largest single global financial market. With trading estimated to be over USD\$1.41 trillion each day, the volume of currency trading is 40 times the level of world equities, including those of emerging financial markets [see White (2000)]. Nearly the entire amount traded is in the three major currencies of the world, the US Dollar, the European Euro and the Japanese Yen.<sup>1</sup>

Many of the concepts introduced in this chapter will be used later in the thesis. Section 2.2 provides a general review of factors which influence exchange rate (price) movements in forex markets. Non-linear times series models aimed at modelling movements in volatility with specific focus on foreign exchange price movements are presented in Section 2.3. These models include the Autoregressive Conditional Heteroscedastic (ARCH) model and its variants. Section 2.4 presents the hypothesis of PPP, which is one of the most conventional approaches to investigate exchange rate movements in the long-term. The chapter

---

<sup>1</sup> McCauley (1997) estimates that, after the introduction of the Euro, 87 percent of all transactions would have one of these three currencies on its side.

then moves on to review modern forex markets in Section 2.5. In Section 2.6 a brief summary is provided to conclude the chapter.

## **2.2 Foreign Exchange Price Movements**

The influences of exchange rate movements are widespread. For instance, the value of the Australian currency is a key factor affecting domestic producers and exporters of minerals and energy commodities.<sup>2</sup> Further, when overseas demand for Australian primary produce and minerals is high, the Australian Dollar appreciates because of capital inflows.

A fixed exchange rate is an exchange rate between the currencies of two countries that is fixed at some level and adjusted only infrequently by central banks. A floating exchange rate is an exchange rate determined by the forces of supply and demand. After World War II, most nations adopted a pegged exchange rate system. Under this system, a government traded enough US Dollars or gold in exchange for its own money to keep a steady exchange rate. Since the early 1970s the major trading countries have generally adopted floating exchange rates which respond to the laws of supply and demand.

---

<sup>2</sup> Australian Commodities (September quarter 1999) published by the Australian Bureau of Agricultural and Resource Economics, claims that increases in earnings of Australian commodity exporters, derived from higher world prices, would be substantially eroded by the expected higher value Australian Dollar.

In general, exchange rates are influenced by a number of fundamental economic factors such as inflation and interest rates, balance of payments, fiscal and monetary policies, and government attitudes to intervention. Currency speculation, wars, elections and market sentiment can also influence currency movements. Exchange rates can move as a result of government intervention to smooth the market. However governments are not able to control the exchange rate over a long period without regard to the fundamentals.

### **2.2.1 Non-linearity in Forex Markets**

The existence of non-linearity in forex markets has been confirmed [Adrangi *et al* (2001), De Gauwe *et al* (1993)]. Heterogeneity both in investors' objectives and in their expectations are often considered to be the major sources of non-linearity in forex markets. Peters (1994) and Guillaume *et al* (1995) show that a non-linear exchange rate process could be generated by heterogeneity in investors' objectives. This type of heterogeneity arises from different investment horizons, geographical locations, institutional constraints and various types of risk profile. It is one explanation why investors can respond differently to the same set of news.

De Gauwe *et al* (1993) show that chaotic and non-linear exchange rate dynamics can be generated by the interaction of fundamentalists and chartists. Brock and Hommer (1996) propose an asset pricing model in which traders move between different beliefs according to the profitability of expectations predictions.



Different complex asset price dynamics can be generated from this structural non-linear model.

While the above studies emphasised heterogeneity, Gouriéroux (1997) stresses the non-linearity in forex markets resulting from market friction, such as the interventions of banks in the markets, the times and size of block trading, and the existence of thin trading.

### **2.3 Volatility Studies**

The volatility of a financial variable is its degree of (random) variability. In financial markets volatility is observed through price movements. However the very nature of financial markets is that gains and losses arise from price movements. In a market in which prices were stable, there would be little incentive to trade. On the other hand if the level of volatility were too high, the market may become unstable. Thus it is argued an appropriate level of volatility is required to maintain a liquid and viable market [Brailsford (1994)].

Volatility was formalised in Markowitz's (1952) mean variance model of portfolio selection. It was later demonstrated to play a crucial role in option pricing in the Black-Scholes (1973) model. Another significant and practical role for volatility is in the calculation of Value at Risk (VaR). VaR has been used for describing capital adequacy standards for banks, and corporate risk exposures to derivatives by investors.

Volatility can be measured in a variety of ways, such as deviations from moving averages or a standard deviation of returns. Since the measure of volatility depends on the theory of asset pricing under test, it is common to use the average of squared daily returns as a volatility measure [see Brailsford and Faff (1996)].

### 2.3.1 ARCH Modelling

The non-linearity in forex markets creates a need to conduct non-linear modelling. ARCH / Generalised ARCH (GARCH) models and their variants have been used to capture non-linearity. Engle (1982) introduces the ARCH models which follow the tradition of the conditional mean models, in particular ARMA models, which are time dependent. The presumption that forecasts of variance at some future time can be improved by using prior information through time dependence is used as a basis for constructing the ARCH models. Friedman and Vandersteel (1982) claim that ARCH modelling can explain the observed volatility clustering of exchange rates and can provide for normal conditional distributions and symmetric - but not leptokurtic - unconditional ones. Large volatility movements are succeeded by further large volatility movements of either sign.

An ARCH (1) model of returns is:

$$y_t = \varepsilon_t \sigma_t \text{ and } \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2, \quad t = 1, \dots, T,$$

where  $\varepsilon_t \sim \text{NID}(0,1)$ ,  $\text{var}(y_t) = \sigma_t^2$ , and NID denotes a normal distribution with independent observations. Also the parameter  $\alpha_1$  has to be non-negative and  $\alpha_0 \geq 0$  to ensure that  $\sigma_t^2 \geq 0$  for all  $t$ .

The GARCH modelling proposed by Bollerslev (1986) is an extension of the ARCH model. Bollerslev *et al* (1992) demonstrate that lower order GARCH specification tends to be more parsimonious but effective in capturing the temporal behaviour of volatility. A GARCH(1,1) model can be expressed as

$$y_t = \varepsilon_t \sigma_t \text{ and } \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad t = 1, \dots, T.$$

The GARCH(1,1) model allows for a longer memory in the conditional variance process than the ARCH model. However the simple GARCH model requires the assumption of symmetric response to shocks. Nelson (1990) proposes the exponential GARCH (EGARCH) model which allows for negative correlation between returns and volatility that is often observed in empirical stock market studies. The EGARCH model is valuable in capturing asymmetric shocks because it defines conditional variance as an asymmetric function of lagged residuals and it does not require non-negativity constraints. The integrated GARCH (IGARCH) model [see Engle and Bollerslev (1986) and Nelson (1990)] is useful in cases when the volatility has a unit root, and a shock to the conditional variance is persistent in the sense that it affects future forecasts at all horizons.

There are many other variants of ARCH models. Some of the more influential ones include:

- Threshold ARCH (TARCH): These models are proposed by Zakoian (1990). TARCH models have different parameters for  $y_{t-1} > 0$  and  $y_{t-1} \leq 0$ . Glosten *et al* (1993) introduce a model of the type:

$$\sigma_t^2 = \alpha_0 + \alpha_1^+ I(y_{t-1} > 0) y_{t-1}^2 + \alpha_1^- I(y_{t-1} \leq 0) y_{t-1}^2,$$

$$\text{where } I(y_{t-1} > 0) = \begin{cases} 1, & y_{t-1} > 0 \\ 0, & y_{t-1} \leq 0 \end{cases} \text{ and } I(y_{t-1} \leq 0) = \begin{cases} 1, & y_{t-1} \leq 0 \\ 0, & y_{t-1} > 0 \end{cases}.$$

Engle and Lee (1992) also use a TARCH model. They allow asymmetry to impact the transitory component of volatility, but not the permanent component.

- Asymmetric power ARCH (A-PARCH): Ding *et al* (1993) introduce A-PARCH models. These models are a class of autoregressive conditional heteroskedastic models, and contain a large number of ARCH and GARCH models. The A-PARCH model has a particular power parameter which makes the conditional variance equation non-linear in the parameters. The A-PARCH models also become the logarithmic GARCH model as the power parameter approaches zero. Hentschel (1995) propose an extended A-PARCH model. This model becomes the EGARCH model as the power parameter approaches zero.

- ARCH-Mean (ARCH-M): Engle *et al* (1987) propose the ARCH-M models in which the conditional variance appeared as an explanatory variable in the

conditional mean. The following equation provides an example of an ARCH-M model:

$$Y_t = X_t b + \delta h_t + \varepsilon_t,$$

where  $\varepsilon_t$  satisfies a GARCH model, and variance  $(\varepsilon_t / \varepsilon_{t-1}) = h_t$ .

The generalised ARCH-Mean (GARCH-M) model can be constructed by incorporating moving average parts into the ARCH-M model.

### **2.3.2 Applications of ARCH Modelling**

The applications of ARCH modelling are versatile. In examining European exchange rate volatility, Friedam and Vandesteel (1982) show that ARCH modelling could explain the observed volatility clustering of exchange rates. Diebold and Pauly (1988) use a bivariate ARCH model and detect a structural shift about the time of the establishment of the European Monetary System (EMS). Bollerslev (1990) employs a multivariate GARCH model and finds reduced volatility and greater coherence for the European exchange rates after March 1979. Vlaar and Palm (1993) support ‘jump-diffusion’ GARCH processes to model realignments of the Exchange Rate Mechanism (ERM). Nieuwland *et al* (1994) conclude that a combined jump-GARCH model with conditional t-distribution innovation is the model that most successfully fits the EMS exchange rate returns. Tsoukalas (1996) demonstrates a multivariate GARCH (1,1) model to show currency stability after the creation of the EMS.

Further, Chionis and MacDonald (1997) employ GARCH models to test volatility, volume and heterogeneity of forex measures of dispersion. Their findings support the usefulness of market microstructure concepts in analysing forex markets. Nieuwland *et al* (1998) examine forex risk premia and find that a GARCH-M specification is often appropriate for the premium. Henry and Summers (2000) use various EARCH modelling and find this type of modelling outperforms the random walk models for the real exchange rates. Parikh and Bailey (1998) use GARCH-M modelling to examine the relationship between macroeconomic fundamentals and nominal exchange rates in a group of 15 European countries and the USA. They conclude that only a small fraction of the conditional volatility of the exchange rates studied can explain variation in macroeconomic fundamentals.

## 2.4 Purchasing Power Parity

The PPP theory states that exchange rates between currencies are in equilibrium when their purchasing power is the same in each of the two countries. Formally the PPP condition can be expressed as follows:

$$\frac{P_t}{P_t^*} = E_t,$$

where  $P_t$  denotes domestic price level,

$P_t^*$  denotes foreign price level, and

$E_t$  denotes units of domestic currency per unit of foreign currency.

The basis for PPP is the "law of one price", which states that homogeneous goods selling in different markets should sell at the same price in the absence of transportation costs, taxes and transaction costs.

Therefore PPP implies that the exchange rate between two countries should equal the ratio of the two countries' price level for a fixed basket of goods and services. When a country's domestic price level is increasing relatively, and inflation is rising relatively, that country's currency must depreciate in order to retain PPP.

There are three assumptions in the above description of PPP. First, the existence of competitive markets for the goods and services in both countries is essential. Second, only tradeable goods are influenced by the law of one price. Fixed goods

such as land and local services are not traded between countries. Third, transaction costs including transportation costs and barriers to trade cannot be significant enough to affect the purchasing power of currencies.

However PPP may not determine exchange rate movements in the short-term. Rather, in the short-term, exchange rate movements are driven by news and expectations, such as announcements about interest rate changes, changes in company profits, changes in capital flows and changes in money supply [see Argy (1992)]. In contrast, when PPP validates its existence, this helps to explain the long-term movements of exchange rates. The economic forces behind PPP should eventually equalize the purchasing power of currencies.

A plethora of theoretical and empirical models have been built around PPP. However empirical tests provide inconclusive evidence of its existence. The classical test for PPP is to regress the nominal exchange rate,  $\ln(E_t)$ , against the ratio of domestic to foreign prices,  $\ln(P_t / P_t^*)$ . Standard Wald statistics are then calculated to test whether the coefficient estimates are consistent with the restrictions embodied in the PPP hypothesis. This approach usually yields results that do not accept the PPP hypothesis [see Roll (1979), Frenkel (1981), and Cumby and Obstfeld (1984)]. A major problem with this approach is that the time-series properties of nominal exchange rates and prices are not specifically taken into consideration. If nominal exchange rates and the ratio of domestic to foreign prices are integrated series, and they usually are, then this test could be biased toward rejecting the null hypothesis of PPP.



To overcome this problem, the theory of cointegration proposed by Engle and Granger (1987) has been widely utilised to test for PPP. A long-term cointegrating relationship between both I(1) variables -  $\ln(E_t)$  and  $\ln(P_t / P_t^*)$  - indicates that  $\beta'X_t = \beta'(\ln(E_t), \ln(P_t / P_t^*))' = \varepsilon_t$ . In this relationship  $\beta$  denotes the cointegrating vector and  $\varepsilon_t$  is stationary. Thus it can be concluded that the necessary condition for PPP exists in the long-term. Therefore the necessary condition for PPP refers to the cointegration between the nominal exchange rate and the price ratio. The necessary and sufficient condition means that these two variables are cointegrated and the cointegrating vector is  $\beta' = (1, -1)$ . To test for the necessary condition for PPP, there are a number of procedures available. Corbae and Ouliaris (1990) and Oh (1996) apply a standard test for unit roots to  $\varepsilon_t$ , or the real exchange rate series [Rogoff (1996)]. If the real exchange rate is found to be stationary, then the necessary condition for PPP is accepted. Otherwise, it can be rejected. Dutt and Ghosh (1995) adopt the Phillips-Hansen Fully Modified Ordinary Least Squares procedure to regress  $\ln(E_t)$  against  $\ln(P_t / P_t^*)$ . The Phillips-Hansen procedure corrects for both endogeneity in the data and asymptotic bias in the coefficient estimates. They then apply the Phillips and Ouliaris (1990) test to determine the order of integration of the residuals from this regression for the necessary condition.

Various studies have used the panel data approach to test for PPP.<sup>3</sup> This approach investigates both cross-sectional and time-series variations. While Pedroni (1995), Wu (1996) and Papell (1997) find evidence of PPP, Pedroni (1996) finds

---

<sup>3</sup> A survey of the PPP literature is provided by Fleissig and Strauss (2000).

additional evidence of panel cointegration, where the rejection of the unit root null was not uniformly strong.

Further studies have been conducted by using panel data to increase the power of augmented Dickey-Fuller (ADF) tests. Using a GLS method, Abuaf and Jorian (1990) and Jorian and Sweeney (1996) support real exchange rate stationarity. Adopting the Levin and Lin (1993) method, Wu (1996), Oh (1996), Papell (1997) and Lothian (1997) show that exchange rates follow a stationary process for OECD economies during the floating exchange rate period. Coakley and Fuertes (1997) employ the IPS test proposed by Im *et al* (1997) and find evidence for real exchange rate stationarity, whereas Canzoneri *et al* (1999) document non-stationarity for real exchange rates.

Another recent approach is to investigate the effect of mean reversion in the context of PPP. Lothian (1998), Siddique and Sweeney (1998) and Koedijk *et al* (1998) find evidence to support mean reversion using panel data. Taylor and Sarno (1998) find that, over the post-Bretton Woods period, real exchange rates exhibit mean reversion. Also Taylor and Sarno show that single-equation ADF tests have low power to reject the unit root null for a mean reverting process. Therefore they suggest a multivariate panel unit root, and find unequivocal evidence of mean reversion in all real exchange rates they examined.

In this thesis the PPP condition is tested for a number of bilateral exchange rates using a different approach. In this approach the necessary condition and the necessary and sufficient condition for PPP are sequentially tested in the

framework of subset VECM with zero coefficients.<sup>4</sup> This thesis utilises the method developed by Penm *et al* (1997) to select the optimal ZNZ patterned VECM involving the nominal exchange rate and the ratio of domestic to foreign prices. The VECM, so determined, then forms the basis for testing the necessary and the necessary and sufficient conditions for PPP.

## **2.5 The Status of Global Forex Markets**

The US Dollar, the Euro and the Japanese Yen are three major international currencies. The international financial system is becoming a three-currency system, dominated by the Euro in Europe, the US Dollar, and the Yen in the Far East and Southeast Asia.

### **2.5.1 Three Major International Currencies**

The US Dollar has historically been the main currency in global forex markets. The Euro has been the official currency in the Euro area after 1 January 1999. The Yen has also been the basis for much forex trading in East Asia.

A BIS report prepared by McCauley (1997) suggests the following estimates of the share of the US Dollar, Euro and Yen in global forex trading. After the introduction of the Euro, 46 percent of all transactions have the US Dollar involved on one side, 25 to 30 percent the Euro, and 13 percent the Yen. In

---

<sup>4</sup>See McFarland *et al* (1994) for the necessary and the necessary and sufficient conditions.

comparison, the Australian Dollar has 1.4 percent. But as Australia is a regional financial centre, the Australian Dollar has significant weight in the East Asian economies.

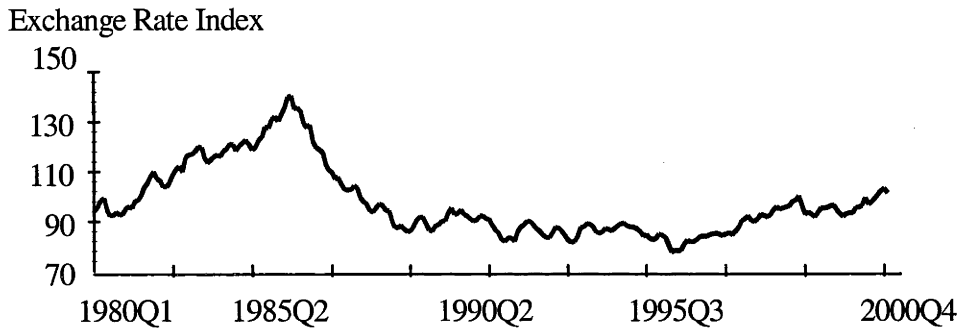
#### **2.5.1.1 The US Dollar**

The US Dollar has historically been the main currency in global forex markets. The Dollar has also been an international reserve and anchor currency due to its well-established stability and size. In particular many countries such as Hong Kong, China and Argentina have adopted *de facto* US Dollar pegged exchange rate regimes to seek price stability, absence of exchange risk, and access to US Dollar trade financing.

As an international currency the Dollar has had a large number of competitive advantages over other currencies. First, it is backed by the world's largest economy and central bank. Second, the US Dollar has a long history of stability. Third, the Dollar is supported by the government of the United States of America, which has a high degree of political stability.

**Figure 2.1**

**Quarterly real effective exchange rate indices of the US Dollar (1980-2000)**



Source: US Federal Reserve

Figure 2.1 shows the quarterly real effective exchange rate indices of the US Dollar relative to the currencies of all other countries in the period 1980 to 2000. The Dollar started from 1980Q1 on an appreciating trend, drifted downwards from 1985Q2, and then gradually rose from 1996Q1. The US economy has been strong over recent years, and this has contributed to a strong US Dollar.<sup>5</sup>

However Figure 2.1 indicates this series appears volatile and non-stationary. Conventional linear models are unlikely to successfully capture the features of this series. This brings a challenge to examine non-linear models, which are needed to capture all features of the series.

---

<sup>5</sup> Although the US Dollar continued its upward trend from 1996 to 2000, the consensus was that, given the absence of an anchor for the US Dollar, its future direction is unpredictable [see Shapiro (1999)].

### **2.5.1.2 The Euro**

With the introduction on 1 January 1999 of the European single currency, the Euro has become the second most widely used currency at the international level, behind the US Dollar and ahead of the Japanese Yen (European Central Bank's Monthly Bulletin – August, 1999).

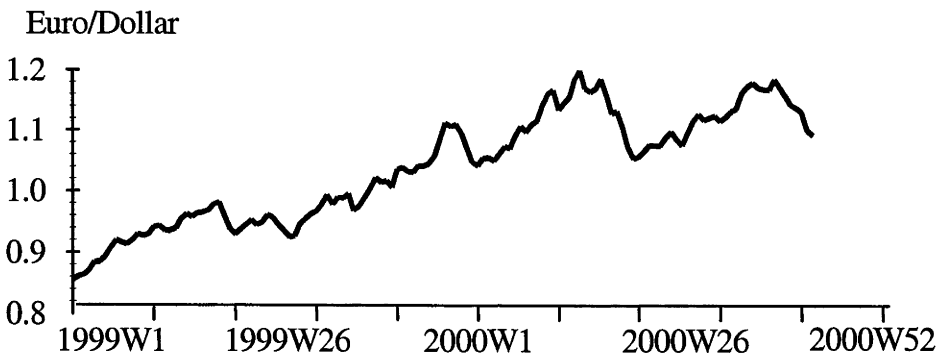
On January 1, 1999 the introduction of the Euro was a significant event in the globalisation of financial markets. The Euro is intended to create broader, deeper and more liquid financial markets in Europe. The main purpose behind the Euro is to improve the price stability and productivity of the European economy. Rather than constant fluctuations in the different exchange rates there will be a more consistent and predictable environment for international trade. Low inflation will protect the value of personal savings and make it easier for both businesses and individuals to plan and invest for the long-term [ECB (1999a)].

The internationalisation of the Euro is illustrated by the increasing popularity of the Euro as a pegging currency. Several Central European countries such as Bulgaria, the Czech Republic, Hungary, Poland, the Slovak Republic and Slovenia have adopted foreign exchange arrangements making use of a basket of currencies in which the Euro is the largest weighting element. Further, in the international corporate bond markets, according to the ECB, the stock of long-term Euro-denominated debt issued by the Euro area governments amounted to around \$2.2 trillion Euro. This market currently stands as the second largest government bond market, only behind the US treasury market.

For the short-term, the European Central Bank needs to establish the Euro's credibility. Since purchasing power per person is even lower in the European Monetary Union (EMU) than in the USA or Japan, it implies a lower nominal interest rate in Europe. Consequently the international demand for the Euro is currently subdued. However the international status of the Euro relies on the expectation that, even though the US Dollar has a prevailing effect at present, international investors will use it.

**Figure 2.2**

**Weekly nominal exchange rates of the Euro relative to the US Dollar (1999-2000)**



Source: DataStream™

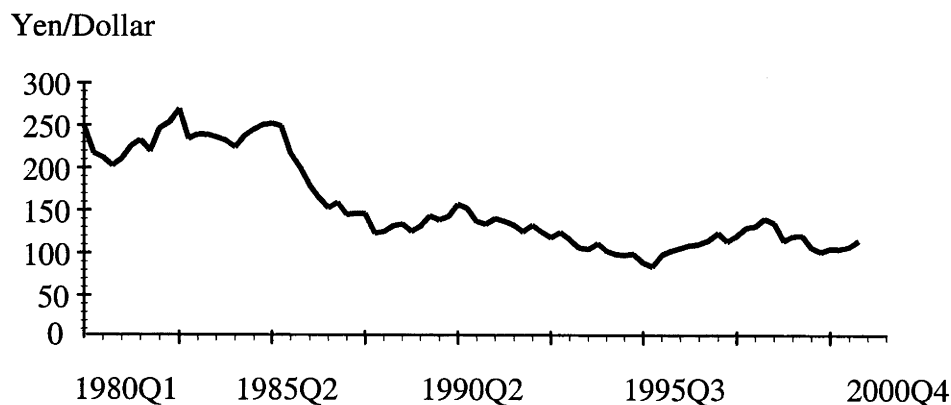
Figure 2.2 shows the weekly nominal exchange rates of the Euro relative to the US Dollar in the period from 1999 to 2000. During this period the Euro depreciated against the Dollar. Although the Euro recovered some ground against the Dollar around June 2000, it weakened again from the third quarter of 2000.

The Euro depreciation against the Dollar is consistent with the strong growth of the US economy compared with slower growth in Europe over the period. Figure 2.2 also indicates fluctuations in the Euro series.

### 2.5.1.3 The Japanese Yen

The Japanese Yen has shown an upward trend to reflect its appreciation relative to the US Dollar since the onset of the Yen's floating in 1971 [see Lothian (1991)]. In late 1996 the Japanese government announced its Big Bang policy. The aim of this policy was to make the Yen, in particular in the Asian loan markets, a truly attractive international currency, challenging the role of the US Dollar.

**Figure 2.3**  
**Quarterly nominal exchange rates of the Japanese Yen relative to the US**  
**Dollar (1980-2000)**



Source: DataStream™



Figure 2.3 shows the quarterly nominal exchange rates of the Japanese Yen relative to the US Dollar in the period from 1980 to 2000. Over the last twenty years, the Yen has strengthened against the Dollar. It has swung substantially during this period, suggesting that the exchange rate series is not stationary.

The Yen has traditionally had a less important role than the US Dollar as an international currency. However continued capital exports by Japan have raised the Yen's international importance in the currency markets. Even before the Asian financial crisis of 1998, many economists recommended against East Asian currencies being pegged to the Dollar. This was because the Dollar-pegged system arguably resulted in overvaluation of East Asian currencies relative to economic fundamentals. From an Asian perspective, Kwan (1994) conducted an analysis of the emerging pattern of trade and interdependence in the Asia-Pacific region, and focused on the implications for output stability in the Asian countries if they pegged their currencies to the Yen. In conclusion, Kwan suggested that the currencies of Hong Kong, Korea, Singapore and Taiwan should peg to the Yen.

However some may suggest there is evidence that the rise of the Yen [see Tse and Ng (1997)], whilst steady, has reflected speculative forces. Although the Japanese inflation and interest rates have been low<sup>6</sup>, the risk that the Yen might be subject to rapid inflation cannot be discounted. Nevertheless, the US Dollar is still in use as the denomination currency in US-Japan trade, and also generally for all

---

<sup>6</sup> Over the period 1998 to 2000 Japan experienced constant or even declining consumer prices. In April 1999 the Bank of Japan announced its intention to maintain zero short-term interest rates until deflation concerns subsided, and the policy was maintained till the end of 2000.

imports of petrol and raw materials into Japan. De Brouwer (2000) has assessed the patterns of common variability in daily changes in various financial markets in both crisis and non-crisis periods in East Asia. De Brouwer's findings indicate that there is little evidence of a 'Yen bloc'.

#### **2.5.1.4 The Australian Dollar**

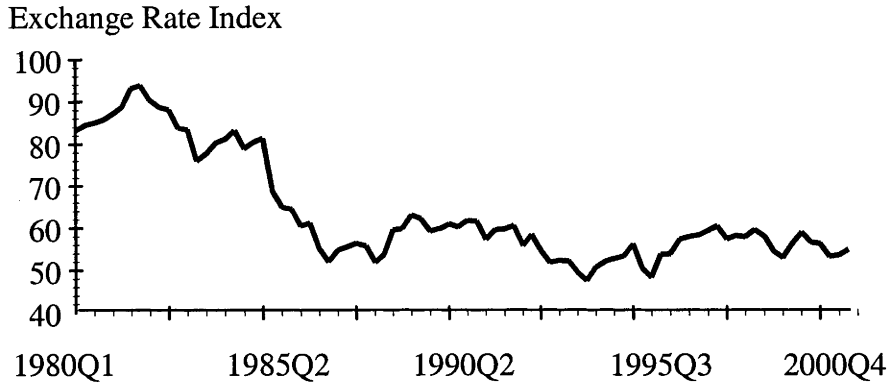
In December 1983 Australia adopted a floating exchange rate. Since then the Australian Dollar has been viewed as one of most flexible exchange rates, with only occasional intervention in the market by the Reserve Bank of Australia (Reserve Bank).<sup>7</sup> The Australian Dollar has become a popular trading currency and has been in demand, because of the volatile nature of trading in the Australian Dollar, and the relatively high nominal interest rates existing on Australian Dollar financial instruments. The Reserve Bank noted that from 1997 Australia was viewed as a proxy for Asian markets and markets of other Asian commodity exporting countries, in which normal market activity had broken down. This added to volatility in the Australian Dollar (see Research Bank of Australia, 1999 March Report).

---

<sup>7</sup> Unless conducting market testing and smoothing very large flows, the Reserve Bank does regularly not intervene in forex markets.

**Figure 2.4**

**Quarterly trade-weighted exchange rate indices of the Australian Dollar  
(1980-2000)**



Source: DataStream™

Figure 2.4 shows the quarterly trade-weighted exchange rate index of the Australian Dollar in the period from 1980 to 2000. The index started from 1980 on an upward trend. However, in December 1983 the Australian Dollar was floated, and the indices later moved in a downward trend in response to market forces. In the year 2000 the Australian Dollar fell by more than 10 percent in trade-weighted terms. From Figure 2.4, it is clear that the series is non-stationary.

This thesis develops ZNZ patterned modelling to seek more sophisticated methodology for capturing the features of exchange rate series. Conducting non-linear modelling in this new framework will provide important insights into modelling and simulation in the analysis of exchange rate series.

### **2.5.1.5 Emerging Financial Markets**

Emerging markets generally include countries facing substantial political, economic, and/or market-specific risks. Interest in Emerging Financial Markets (EFMs) has grown over the past decade. Given political and economic structures that previously existed, often little was known about these markets and international investment levels were low, in part due to high costs of entry. However since the 1990s there have been substantial changes in political and economic environments in many regions such as China, Eastern Europe, Latin America and Russia. As a result emerging markets now represent a feasible investment alternative for international investors, and the last decade has witnessed massive capital flows in and out of EFMs.<sup>8</sup>

Although emerging markets share a potential for big gains, they still present a variety of risks. The broad risks comprise the possibility of government instability, a prolonged recession, a surge in inflation, currency devaluation, political risk and other economic uncertainties. In the financial market the risks include volatile performance, insider trading activity, thin trading activity, dubious investor protection, unclear accounting practices, and inadequate disclosure of crossholdings by majority shareholders. The adverse consequences of these structural distortions may help us understand some of the causes of the recent financial crises in Latin America and East Asia.

---

<sup>8</sup> To illustrate, the International Finance Corporation (IFC, 1995) recorded the aggregate market value of EFMs in 1994 as US\$1,930 billion, up nearly 21 per cent from 1993. New capital raised in 1994 for these markets was US\$51.4 billion.

Traditionally in order to achieve low inflation, many countries in emerging markets have adopted fixed exchange rates. This fixed exchange rate policy has helped many Eastern European and Latin economies to successfully deliver low inflation and a stable currency. This policy also played a significant role in reducing high inflation levels in the East Asian economies before the East Asian currency crisis. However this policy led to a low external competitiveness and a high current account deficit that eventually became untenable. The majority of countries are moving towards true floating exchange rates [see BIS (2001)].

### **2.5.2 The East Asian Currency Crisis**

One of the causes of the East Asian currency crisis is that many East Asian economies initially adopted *de facto* US Dollar pegged exchange rate regimes. Thus, such countries experiencing recessions and sharply depreciating currencies, had to confront an appreciating US Dollar. Thailand has been on a currency basket system since 1984, which required the Bank of Thailand to stabilise the Baht in relation to a basket of foreign currencies in which the weight of the Dollar was large. Many other East Asian economies, including the other ASEAN countries as well as Taiwan and Korea, also adopted either a US Dollar peg system or a currency basket system with a large weight on the Dollar.

In the middle of 1997 the East Asian currency crisis was triggered by speculative attacks on the over-valued East Asian currencies, including the Korean Won,

Indonesian Rupiah and Thai Baht [see Chang *et al* (1997), Edison *et al* (1998)].

According to Stanley Fischer (1988), the major factors leading to the crisis were:

“First, the failure to dampen overheating pressures that had become increasingly evident in Thailand and many other countries in the region and were manifested in large external deficits and property and stock market bubbles; second, the maintenance of pegged exchange rate regimes for too long, which encouraged external borrowing and led to excessive exposure to foreign exchange risk in both the financial and corporate sectors; and third, lax prudential rules and financial oversight, which led to a sharp deterioration in the quality of banks' loan portfolios.”

As a result of the crisis, in late 1997 and through 1998 the Asian crisis countries experienced net capital outflows close to US\$100 billion. Most of the outflows reached both Europe and the United States [see Wincoop and Yi (2000)]. Also this crisis had a very substantial impact on the forex markets of the Asian crisis countries. By mid-July 1998 the Indonesian Rupiah had suffered the largest depreciation of more than 75 percent. The Korean Won, the Thai Baht, the Malay Ringgit and the Philippine Peso had depreciated in a range of 45 to 55 percent (source: DataStream™).

## **2.6 Summary**

This chapter has presented a discussion of price movements in forex markets with emphasis on the recent price movements. It has introduced factors affecting forex market movements and non-linearity in forex price movements, summarised approaches other researchers have adopted in modelling forex volatility, and discussed the PPP hypothesis.

## **CHAPTER 3**

### **NEW MODELLING TECHNIQUES IN FINANCIAL TIME-SERIES**

#### **ANALYSIS**

##### **3.1 Introduction**

Vector time-series modelling is used in this thesis to examine the behaviour of international financial markets, in particular foreign exchange market movements. This chapter presents a range of new techniques including the ZNZ patterned Vector Autoregressive (VAR) modelling, ZNZ patterned Vector Error-Correction (VECM) modelling and ZNZ patterned polynomial neural networks. If key linear and non-linear interactions among variables are accurately captured, these new techniques can then reveal important information about the structure and dynamics of the markets. These techniques can also provide insights into, and improvements in the performance of modelling and simulation methodology for, price movements.

The chapter is organised as follows. Section 3.2 describes VAR modelling. Section 3.3 provides the detailed background to, and considers the use of, the Yule-Walker relations for fitting VAR models. Section 3.4 assesses the problems and theoretical inconsistency that arise by using a two-variable VAR example. Section 3.5 presents an adjustment to the Yule-Walker relations and contains the main theoretical contribution of this chapter. Section 3.6 introduces ZNZ



patterned VECM modelling. Section 3.7 outlines ZNZ patterned polynomial neural networks. Section 3.8 concludes the chapter.

### 3.2 VAR Modelling

VAR models represent an advance in the analysis of vector time-series. In recent years the use of VAR modelling as a means of analysing financial time-series has become common because it is both simple and generalisable. In particular, VAR modelling has been increasingly employed to examine dynamic relationships in both exchange rate and stock markets. VAR modelling also provides a device that has proved to be a more computationally efficient tool, and therefore less costly, than conventional financial and econometric time-series techniques as a means of producing forecasting. Further, VAR modelling can be used for statistical inference such as Granger-causality, impulse response and variance decomposition analysis.

However VAR modelling was originally introduced to avoid the need to incorporate *a priori* highly uncertain restrictions on the structure of models. As a result heavy parameterisation of VAR models has become a major deficiency<sup>9</sup> [see Terrell (1988)]. As the number of parameters to be estimated grows very rapidly, the degrees of freedom will be heavily reduced. Consequently the model's forecasting ability, particularly outside the sample, will be severely

---

<sup>9</sup> Of note, VAR models are not equivalent to simultaneous-equation models, although VAR models can be considered as reduced-form equations.

inhibited. This creates a need to develop ZNZ patterned VAR modelling, with allowance for zero entries in coefficient matrices.

### **3.2.1 VAR Applications**

VAR models have been used widely in financial and economic applications. Eun and Shim (1989) estimate a VAR using index returns on nine stock markets to examine interactions among the markets. In the context of emerging stock markets, Bekaert *et al* (1999) estimate a VAR using capital flows, equity returns, dividend yields and interest rates to examine the extent to which lower interest rates contribute to increased capital flows. In a similar study, Froot *et al* (1998) employ VAR estimation to examine the relationship between capital flows and equity returns in emerging markets.

Kamas (1995) utilises VAR modelling to examine the sources of inflation in several emerging markets. In the case of Colombia, Kamas concludes that changes in domestic credit affect the balance of payments but not the exchange rate. Montiel (1989) uses VAR modelling in explaining the acceleration of inflation in Brazil, Argentina and Israel. His findings indicate that exchange rate movements are the driving force behind the most recent inflation in the first two countries. In the case of Israel, changes in the role of both wage rates and the exchange rate accounts for the surge of inflation. Reinhart and Reinhart (1991) employ a VAR to examine money neutrality for Colombia. The outcome shows that money supply has a strong effect on exchange rate variability, while the exchange rate influences inflation. Dornbusch *et al* (1990) apply VAR modelling

to assess sources of inflation in Argentina, Brazil, Mexico and Peru. The results suggest that the exchange rate shocks have a significant impact on inflation.

Chadha and Hudson (1998) conduct VAR estimation of a model involving real exchange rates, output and prices for 14 European Union (EU) countries to examine the optimum currency case for the European Monetary Union. McNelis and Asilis (1995) utilise VAR estimation to model exchange rates, the Nikkei index, the Dow-Jones Industrial average and six other relevant variables to compare the different outcomes, from Nash and cooperative equilibria, for the United States and Japan. Kim *et al* (2000) provide VAR estimation in the relevant exchange rate, the US market index and the price of the underlying shares in the local currency, to investigate price transmission dynamics between American Depository Receipts and their underlying foreign securities. Neely and Weller (2000) carry out VAR estimation in six variables that are believed to affect stock prices for each of three two-country cases in order to evaluate predictability in international asset returns. They emphasise the importance of the stability of VAR modelling for out-of-sample predictability.

Further examples of financial and economic variables that have been tested for Granger-causal relations by using the conventional VAR modelling include volatility transmission [Bhattacharya *et al* (2000)], stock markets and foreign exchange [Bekaert and Hodrick (1992)], volatility of stock market returns [Whitelaw (1994)], monetary policy and the stock market [Bernanke and Blinder (1992), Thorbecke (1997)], cross-market relationships [Malliaris and Urrutia (1992)] and interest rates [Hassapis *et al* (1999)].

### 3.2.2 ZNZ Patterned VAR Modelling

The use of VAR models for financial and economic research has in part been driven by the desire to provide users with a relatively simple modelling and forecasting procedure. However early researchers realised that heavy parameterisation of their VAR models resulted in poor *ex-ante* forecasting performance. Their proposed procedures rest on the assumption that the coefficient matrices of the VAR model have all non-zero entries. In practice the assumed specification for this full order VAR model can be quite different from the actual specification, as there can be zero entries in the coefficient matrices of the model. In effect, the assumption of non-zero entries restricts the range of possible model specifications. Further, if the true underlying VAR process has zero entries in its structure, then sub-optimal model design induced by assuming a full-order structure can produce misleading inferences and inferior projections. Consequently, models have been developed that allow for zero entries in the coefficient matrices such as a zero-non-zero, ZNZ, patterned structure. However, implementation of a ZNZ structure in a VAR is difficult given the large number of parameters and possibilities. That is, in the absence of an effective approach to find the optimal model, relaxation of the assumption of non-zero entries is problematic.

The issue is also relevant when investigating causality. Optimal VAR models with ZNZ patterned coefficient matrices can also be used as a basis for detecting Granger-causality and instantaneous causality among time-series variables. The

most successful applications in ZNZ patterned VAR modelling are associated with Granger-causality, Granger non-causality and indirect causality detection. This is because both Granger non-causality and indirect causality detection are crucially dependent on making use of zero coefficient entries in the true structure, where the structure does indeed include several additional zero entries to those representing causal structures. Application of VARs to economic and financial time-series data has revealed that zero entries are indeed possible [see Caines *et al* (1981) and Penm *et al* (1992, 1999)]. Since the ZNZ patterned VAR modelling allows for zero entries, the results of simulations and applications of methodology [see Penm and Terrell (1984a, 1984b), Brailsford *et al* (2001a, 2001d)] have indicated that the selected optimal ZNZ patterned VAR provides a straightforward, certain and effective means of indicating all Granger-causality, Granger non-causality and indirect causality from the coefficient matrices on the lagged terms.

One approach to select the optimal ZNZ patterned VAR model has been advocated by Penm and Terrell (1984a), and it centres on their development of a search algorithm using the Yule-Walker relations in conjunction with model selection criteria. However, that approach does not examine the estimation of the residual variance-covariance relation, rather it focuses only on the Yule-Walker coefficient relations. A direct extension of the use of the Yule-Walker relations for fitting vector autoregressive models where ZNZ patterned coefficient matrices are present is inconsistent with statistical procedures as the resultant estimated variance-covariance matrix of the white noise disturbance process becomes non-symmetric. In Section 3.5, an approach is provided that considers a consistent

adjustment to the variance-covariance relation within the Yule-Walker relationship and leads to an effective approach to identify the optimal ZNZ model within the context of a VAR system. The development of the approach in this chapter is a significant contribution given the extant literature that seeks to employ full order estimation.

### 3.3 Using the Yule-Walker Relations for Fitting VAR Models

In this section, the fitting of a VAR with zero coefficient restrictions is presented.

First, let  $y(t) = \{y_1(t), y_2(t), \dots, y_s(t)\}'$  be a zero mean, wide-sense stationary time-series of dimension  $s$ . Consider the vector AR ( $p$ ) model of the form:

$$\sum_{k=0}^p A_k y(t-k) = A^p(L)y(t) = \varepsilon(t), \quad (3.1)$$

where  $A_0=I$ ,  $A_k$ ,  $k=1, \dots, p$  are the  $s \times s$  parameter matrices, and

$$A^p(L) = I + \sum_{k=1}^p A_k L^k.$$

$L$  denotes the lag operator, and the roots of  $|A^p(L)| = 0$  lie outside or on the unit circle. Also  $\varepsilon(t)$  is an  $s \times 1$  stationary vector process with  $E\{\varepsilon(t)\} = 0$ , and thus:

$$E\{\varepsilon(t)\varepsilon'(t-k)\} = \begin{cases} V & \text{as } k=0 \\ 0 & \text{as } k>0 \end{cases}. \quad (3.2)$$

The sample lag covariance matrices,

$$\Gamma_k = \frac{1}{N} \sum_{t=1}^{N-k} y(t)y'(t+k), \quad (3.3)$$

obey the following Yule-Walker relations.

The Yule-Walker coefficient relations are:

$$\Gamma_j + \sum_{k=1}^p \hat{A}_k \Gamma_{j-k} = 0 \quad (j=1, \dots, p). \quad (3.4)$$

The Yule-Walker residual variance-covariance relation is:

$$\Gamma_0 + \sum_{k=1}^p \hat{A}_k \Gamma_{-k} = \hat{V}, \quad (3.5)$$

where  $\Gamma_k = \Gamma'_{-k}$ ;  $N$  is the sample size,  $\hat{A}_k$  and  $\hat{V}$  are the estimates of  $A_k$  and  $V$  respectively, and  $|\hat{V}|$  is described as the generalised residual sum of squares.

In a full-order VAR model, possible models with zero coefficient elements are neglected. For example in a bivariate VAR model when  $p = 5$ , the coefficients,  $A_1$ ,  $A_2$ , up to and including  $A_5$  are assumed non-zero. However there are  $2^{20} = 65,536$  possible models in this example. Thus a large number of possible models will be ignored under the assumptions of all non-zero coefficients. More important, if the true underlying VAR process has a ZNZ patterned structure, a sub-optimal model

design from an imposed full-order structure can produce misleading inferences and inferior projections.

Penm and Terrell (1984a) propose a search algorithm, using the Yule-Walker relations for fitting VAR models in conjunction with model selection criteria, to select the optimal ZNZ patterned VAR models. Background information on the fitting of VAR models using the Yule-Walker relations is presented in Section A.1 of Appendix A. In the course of using the Yule-Walker relations to conduct the fitting of ZNZ patterned VAR (p) models, only the following p+1 lag covariance matrices are required to compute the estimated coefficient matrices and residual variance-covariance matrix:

$$\Gamma_0, \Gamma_1, \dots, \Gamma_p.$$

However, the estimated V using the usual least squares (LS) method is as follows:

$$\hat{V} = \frac{1}{N-p} \sum_{i=p+1}^N \hat{\varepsilon}_i \hat{\varepsilon}_i',$$

where  $\hat{\varepsilon}_i$  denotes the estimate of  $\varepsilon(i)$ .



This method suffers from the need to estimate and store all individual  $m \times 1$  residual vectors,  $\hat{\epsilon}_t$ ,  $t=1,2,\dots,N$ , and then compute  $\hat{V}$  for each ZNZ patterned VAR model. In order to estimate individual residual vectors, all observation vectors  $y(t)$ ,  $t=1,2,\dots,N$ , must be held in storage to carry out  $\hat{\epsilon}(t)$  estimation. When using the LS method, a very large number of candidate ZNZ patterned VAR models must be estimated before the optimal model can be selected, which involves a considerable amount of computational cost in terms of execution time and memory storage, and these costs are important considerations.

Since the LS method requires the storage of all observation vectors and then the use of these vectors to estimate individual residual vectors, the need for heavy computing resources becomes a heavy burden for large sample cases. It is obvious that estimation of the residual variance-covariance matrix, which minimizes the need for computing resources, becomes an important issue. As outlined later in the chapter, there is no need to estimate individual residual vectors when an adjustment is made to the Yule-Walker approach. Hence this approach is simple and avoids considerable computational costs.

The issue in Penm and Terrell (1984a) is that their estimate of  $V$  using the Yule-Walker residual variance-covariance relation of (3.5) is not analyzed. Only the Yule-Walker coefficient relations in (3.4) are canvassed. A direct extension of the Yule-Walker residual variance-covariance relation to fit the ZNZ patterned VAR model is inappropriate as it is inconsistent with statistical theory. The problem arises because the resultant estimated variance-covariance matrix of the white noise process becomes non-symmetric, violating the condition that  $V$  must be

symmetric. This violation has important implications. One consequence is that VAR cannot be converted to an equivalent vector moving average (VMA) model as proposed in Penm and Terrell (1986, 1994) to conduct testing for Granger-causality. Further, the innovation accounting proposed by Lee (1992) will not work under these conditions (see Section A.2 of Appendix A). Hence, this failure to ensure symmetry of estimates of  $V$  creates the motivation for developing an adjustment in this chapter to the Yule-Walker relations for fitting of ZNZ patterned VAR models.

An alternative solution is to use other approaches that do not rely on the Yule-Walker relations. However, each of these approaches is problematic, particularly in terms of their large computational costs. The alternative methods are briefly outlined below.

### 3.3.1 Alternative Approaches

First, consider the standard least squares (LS) approach. As described in Section A.1 of Appendix A, for fitting a full-order VAR ( $p$ ) model using the Yule-Walker relations, the following block Toeplitz matrix  $C_{p+1}$  can be constructed:

$$C_{p+1} = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_p \\ \Gamma_{-1} & \Gamma_0 & \cdots & \Gamma_{p-1} \\ \vdots & \vdots & & \vdots \\ \Gamma_{-p} & \Gamma_{1-p} & \cdots & \Gamma_0 \end{bmatrix} = \begin{bmatrix} C_p & \vdots \\ \cdots & \Gamma_0 \end{bmatrix}, \quad (3.6)$$

and the following relation for the estimate of  $|V|$ :

$$|\hat{V}_p| = |C_{p+1}| / |C_p| = |d_{p+1}|, \quad (3.7)$$

where  $d_i$ ,  $i = 1, \dots, p+1$  are diagonal block entries of the block diagonal matrix resulting from a block Choleski decomposition for  $C_{p+1}$ . This outcome indicates that in the course of computing  $|\hat{V}_p|$  for the VAR (p) model, the generalised residual sums of squares for all the lower order VAR models fitted to the data are also obtained.<sup>10</sup> However as described in Section A.1 of Appendix A, this outcome cannot be achieved by using the conventional LS approach. Since  $R_{LS}(p)$  as defined in (A.9) for each different VAR model must be reconstructed from the observations to conduct individual fittings, and the observations must be saved in storage for reconstructing  $R_{LS}(i)$ ,  $i = 1, \dots, p$ , a considerable amount of computational cost in terms of execution time and data storage will be generated. Note that these weaknesses of the conventional LS method also exist in the remaining steps of selecting the optimal VAR, and become severe when the number of lags, or the number of variables, in the system of (3.1) is large. Thus the commonly employed LS approach is considerably more computationally costly than the Yule-Walker approach.

Second, the generalized least squares (GLS) method can be conducted by applying the conventional LS approach as a basis. After the symmetric and positive definite  $\hat{V}$  is estimated by the LS method, there exists an  $m \times m$  non-singular matrix  $\hat{K}$ , such that  $\hat{V}^{-1} = \hat{K}\hat{K}'$ . The method pre-multiplies  $y(t)$  by  $\hat{K}^{-1}$ , and then follows the LS estimation for fitting of the VAR models to obtain the

conventional GLS estimates. However as the LS approach to conduct the selection of the optimal ZNZ patterned VAR is computationally expensive when the number of possible candidate models could be billions, the conventional GLS method will similarly suffer from excessive computational costs.

Third, the maximum likelihood (ML) approach is a non-linear approach but becomes infeasible whenever the number of parameters is very large [Chen and Zadrozny (1998)]. In addition there exist innumerable candidate models in the ZNZ patterned VAR environment. The ML approach needs to apply to each individual VAR model independently, and no previous computational information can be utilised.

Chen and Zadrozny (1998) propose the extended Yule-Walker equation to estimate a VAR for mixed frequency data. The estimated  $V$  for their approach is as follows:<sup>11</sup>

$$\hat{V} = \frac{1}{N-p} \sum_{i=p+1}^N \hat{\epsilon}_i \hat{\epsilon}_i',$$

which is identical to the conventional LS approach. Thus the approach of Chen and Zadrozny also needs to consider each VAR model independently for estimation of the individual residual variance-covariance matrices. In complete data cases (ie. no missing values), their approach only concerns full-order models.

---

<sup>10</sup> The proposed model selection criteria use the generalised residual sums of squares.

<sup>11</sup> See Section 3 in Chen and Zadrozny (1998).

The ZNZ patterned modelling with no missing data is not investigated in Chen and Zadrozny.<sup>12</sup>

Section A.1 of Appendix A shows that the conventional LS method is quite different from the Yule-Walker approach. Thus,  $\hat{V}$  using the LS method is also quite different from  $\hat{V}$  under the Yule-Walker approach. It follows that the approach of Chen and Zadrozny has ignored the issue of estimating the residual variance-covariance matrix. Although  $\hat{V}$  using the LS method is asymptotically equivalent to  $\hat{V}$  using the Yule-Walker approach, these two estimators can be quite different in a finite sample. If  $\hat{V}$  proposed in Chen and Zadrozny is estimated using the Yule-Walker approach, then in the case of complete data the approach in this chapter can be employed to select the optimal ZNZ patterned VAR. Thus, again a considerable amount of computational costs can be avoided.

The most successful applications in ZNZ patterned VAR modelling are associated with Granger non-causality and indirect causality detection. This is because both Granger non-causality and indirect causality detection are crucially dependent on making use of zero coefficient entries in the true structure, where the structure does indeed include several zero entries. Application of VARs to economic and financial time-series data has revealed that zero entries are indeed possible [Caines *et al* (1981), Penm *et al* (1992)]. Since the ZNZ patterned VAR modelling allows for zero entries, the selected optimal ZNZ patterned VAR

---

<sup>12</sup> However the approach in Chen and Zadrozny (1998) addresses an interesting topic of estimation for mixed frequency data. Incorporating their approach into the ZNZ patterned modelling for mixed frequency data warrants further investigation.

provides a straightforward, certain and effective means of indicating the presence of Granger-causality, Granger non-causality and indirect causality from the coefficient matrices on the lagged terms.

### 3.4 The Inconsistency in Using Yule-Walker Relations

This section shows the theoretical inconsistency of the use of the Yule-Walker relations for fitting of ZNZ patterned VAR models using a two-asset example.

In considering a ZNZ patterned VAR model, zero entries are allowed for in the parameter matrices  $A_k$  of (3.1). Let the returns of the assets,  $\Delta y_{1,t}$ ,  $\Delta y_{2,t}$ , be jointly determined by the following two equations:

$$\Delta y_{1,t} + a_{12}\Delta y_{2,t-1} = \varepsilon_{1,t} \quad (3.8)$$

$$\Delta y_{2,t} + a_{21}\Delta y_{1,t-1} + a_{22}\Delta y_{2,t-1} = \varepsilon_{2,t}, \quad (3.9)$$

where  $y_{1,t}$  and  $y_{2,t}$  are the log prices of the assets. In this two-equation system the first equation shows that  $y_{1,t}$  is caused by  $y_{2,t}$ , while the second equation indicates that  $y_{2,t}$  is caused by  $y_{1,t}$ . Thus a feedback relation exists between these two asset prices.

The equivalent VAR model of this system can then be expressed as:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} + \begin{bmatrix} 0 & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad (3.10)$$

where the white noise process comprises two components  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$ , with:

$$E\{\varepsilon_{1,t}\} = E\{\varepsilon_{2,t}\} = 0, \text{ and}$$

$$E\left\{ \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-k} & \varepsilon_{2,t-k} \end{bmatrix} \right\} = \begin{cases} V & \text{as } k = 0 \\ 0 & \text{as } k > 0. \end{cases}$$

This section now uses the Yule-Walker coefficient relations to estimate  $a_{12}$ ,  $a_{21}$  and  $a_{22}$ . Since  $a_{11} = 0$ , and  $\varepsilon_{1,t}$  is uncorrelated with the asset return,  $\Delta y_{2,t-1}$ , the following relation derived from the first equation of (3.10) is apparent:

$$E\{\Delta y_{1,t} \Delta y_{2,t-1}\} + \hat{a}_{12} E\{\Delta y_{2,t-1} \Delta y_{2,t-1}\} = 0,$$

where  $\hat{a}_{ij}$  are the estimates of  $a_{ij}$ . Thus the estimate  $\hat{a}_{12} = -\frac{\tau_{12}(1)}{\tau_{22}(0)}$  will be

achieved, where the correlation functions between asset returns,

$$\tau_{ij}(k) = \frac{1}{N} \sum_{t=1}^{N-k} \Delta y_i(t+k) \Delta y_j(t) = \frac{1}{N} \sum_{t=1}^{N-k} \Delta y_j(t) \Delta y_i(t+k) = \tau_{ji}(-k).$$

From (3.10), since the asset return vector  $\begin{bmatrix} \Delta y_{1,t-1} & \Delta y_{2,t-1} \end{bmatrix}$  is uncorrelated with

$\begin{bmatrix} \varepsilon_{1,t} & \varepsilon_{2,t} \end{bmatrix}$ , the following arises:

$$\begin{bmatrix} \tau_{11}(0) & \tau_{21}(0) \\ \tau_{12}(0) & \tau_{22}(0) \end{bmatrix} \begin{bmatrix} \hat{a}_{21} \\ \hat{a}_{22} \end{bmatrix} = - \begin{bmatrix} \tau_{21}(1) \\ \tau_{22}(1) \end{bmatrix}. \quad (3.11)$$

Hence,

$$\hat{a}_{21} = \frac{\tau_{21}(0)\tau_{22}(1) - \tau_{22}(0)\tau_{21}(1)}{\tau_{11}(0)\tau_{22}(0) - \tau_{21}(0)\tau_{12}(0)} \text{ and } \hat{a}_{22} = \frac{\tau_{12}(0)\tau_{21}(1) - \tau_{11}(0)\tau_{22}(1)}{\tau_{11}(0)\tau_{22}(0) - \tau_{21}(0)\tau_{12}(0)}.$$

Thus the coefficient estimates in terms of the correlation functions between asset returns are established. Of note, the use of the above approach is identical to the use of equation (A.11a) as proposed in Appendix A for fitting of the ZNZ patterned VAR models.

As a result the estimate  $\hat{V}$  in equation (3.5) becomes:

$$\begin{bmatrix} \tau_{11}(0) & \tau_{12}(0) \\ \tau_{21}(0) & \tau_{22}(0) \end{bmatrix} + \begin{bmatrix} 0 & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \end{bmatrix} \begin{bmatrix} \tau_{11}(-1) & \tau_{12}(-1) \\ \tau_{21}(-1) & \tau_{22}(-1) \end{bmatrix}.$$

Because  $\tau_{ij}(k) = \tau_{ji}(-k)$ ,  $\hat{V}$  can be expressed as:

$$\begin{bmatrix} \tau_{11}(0) & \tau_{12}(0) \\ \tau_{12}(0) & \tau_{22}(0) \end{bmatrix} + \begin{bmatrix} \hat{a}_{12}\tau_{12}(1) & \hat{a}_{12}\tau_{22}(1) \\ \hat{a}_{21}\tau_{11}(1) + \hat{a}_{22}\tau_{12}(1) & \hat{a}_{21}\tau_{21}(1) + \hat{a}_{22}\tau_{22}(1) \end{bmatrix}, \quad (3.12)$$

which is non-symmetric. Intuitively,  $V$  is symmetric in the true model of (3.1) and there is a need for the estimate  $\hat{V}$  to conform to the behaviour of  $V$ . Therefore the estimate  $\hat{V}$  must be a symmetric matrix. As described earlier, this non-symmetric  $\hat{V}$  violates the symmetric condition required in Lee (1992) and in



Penm and Terrell (1986). This violation indicates that, in practice, the innovation accounting described in Lee will not work (see Section A.2 of Appendix A), and a VAR model cannot be converted to its equivalent VMA model as proposed in Penm and Terrell to conduct testing for Granger-causality. Thus an adjustment to the Yule-Walker relations is required.

### 3.5 The Adjustment to the Yule-Walker Relations

The necessary adjustment to the Yule-Walker relations for fitting of VAR models with ZNZ patterned coefficient matrices follows directly from the inconsistency demonstrated in the previous section.

With the definition of the variance-covariance matrix in (3.2),

$$\begin{aligned} V &= E \left\{ \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix} \right\} \\ &= E \left\{ \left\{ \begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} + \begin{bmatrix} 0 & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \end{bmatrix} \right\} \left[ \Delta y_{1,t} \quad \Delta y_{2,t} \right] + \begin{bmatrix} \Delta y_{1,t-1} & \Delta y_{2,t-1} \end{bmatrix} \begin{bmatrix} 0 & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \right\} \right\}. \end{aligned}$$

Using  $\tau_{ij}(k) = \tau_{ji}(-k)$ , the estimate  $\hat{V} = \begin{bmatrix} \tau_{11}(0) & \tau_{12}(0) \\ \tau_{12}(0) & \tau_{22}(0) \end{bmatrix} +$

$$\begin{aligned} & \begin{bmatrix} \hat{a}_{12}\tau_{12}(1) & \hat{a}_{12}\tau_{22}(1) \\ \hat{a}_{21}\tau_{11}(1) + \hat{a}_{22}\tau_{12}(1) & \hat{a}_{21}\tau_{21}(1) + \hat{a}_{22}\tau_{22}(1) \end{bmatrix} + \begin{bmatrix} \hat{a}_{12}\tau_{12}(1) & \hat{a}_{21}\tau_{11}(1) + \hat{a}_{22}\tau_{12}(1) \\ \hat{a}_{12}\tau_{22}(1) & \hat{a}_{21}\tau_{21}(1) + \hat{a}_{22}\tau_{22}(1) \end{bmatrix} + \\ & \begin{bmatrix} 0 & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \end{bmatrix} \begin{bmatrix} \tau_{11}(0) & \tau_{12}(0) \\ \tau_{12}(0) & \tau_{22}(0) \end{bmatrix} \begin{bmatrix} 0 & \hat{a}_{21} \\ \hat{a}_{12} & \hat{a}_{22} \end{bmatrix} \end{aligned} \quad (3.13)$$

Since the first matrix of equation (3.13) is symmetric, the second matrix is the transpose of the third matrix, and the remaining product matrix is also symmetric, therefore the matrix  $\hat{V}$  is symmetric.

An analogous approach to equation (3.13) is feasible. From equation (3.2),

$$V = E \left\{ \left[ \sum_{k=0}^p A_k y(t-k) \right] \left[ \sum_{j=0}^p y'(t-j) A_j' \right] \right\}.$$

Then,

$$\hat{V} = \Gamma_0 + \sum_{k=1}^p \hat{A}_k \Gamma_{-k} + \sum_{j=1}^p \Gamma_j \hat{A}_j' + \sum_{j=1}^p \sum_{k=1}^p \hat{A}_k \Gamma_{j-k} \hat{A}_j'. \quad (3.14)$$

It is obvious that this  $\hat{V}$  is symmetric. Since  $\Gamma_{-j}' = \Gamma_j$ ,  $\Gamma_0$  is symmetric. If  $k$  is redefined as  $j$ , the second matrix is the transpose of the third matrix. If  $j$  and  $k$  are redefined as  $k$  and  $j$  respectively, the fourth matrix becomes  $\sum_{k=1}^p \sum_{j=1}^p \hat{A}_j \Gamma_{k-j} \hat{A}_k'$ , which is the transpose of the fourth matrix itself.

In addition, comparing equation (3.14) to the estimator of  $V$  using individual residual vectors, the structure of (3.14) is computationally efficient in terms of execution time and storage requirements, and provides the obvious relations to link the covariance matrices with different lags.

Of note, consideration of the contemporaneous correlation in  $\varepsilon(t)$  cannot be ignored. A ZNZ patterned VAR model can be viewed as a system of ‘seemingly unrelated regressions’ as originally proposed by Zellner (1962). As the regressors in each equation of the VAR model are no longer necessarily the same, the generalised least squares (GLS) coefficient estimator using the Yule-Walker relations for the ZNZ patterned VAR is more efficient than the estimator using equation (3.14). Brailsford *et al* (2001d) show that this GLS estimator is an asymptotic approximation to the ML estimator. Henceforth the notation GLS-YW is used for this estimator. As described earlier,  $\hat{V}^{-1} = \hat{K}\hat{K}'$ .  $y(t)$  is premultiplied by  $\hat{K}^{-1}$ . Then the proposed method of using the Yule-Walker relations for fitting of VAR models is followed so the GLS-YW coefficient estimates of the ZNZ patterned VAR model can be obtained.

### 3.6 VECM Modelling

Recent cointegration work proposed by Engle and Granger (1987) has suggested that if a time-series system under study includes cointegrated variables, then this system may be more appropriately specified as a vector error-correction model (VECM) rather than a VAR [see Engle and Granger (1987)]. The VECM is identical to the VAR model with unit roots, as the following indicates.

In VAR modelling of (3.1), the following identity emerges:

$$A^p(L) = A^p(1) + (I - L)(I + \sum_{\tau=1}^{p-1} A^*_\tau L^\tau). \quad (3.15)$$

In accordance with the concept of cointegrated variables introduced by Granger (1981),  $y(t)$  is said to be  $I(1)$  if it contains at least one element which must be differenced before it becomes  $I(0)$ . Then  $y(t)$  is said to be cointegrated of order 1 with the cointegrating vector,  $\beta$ , if  $\beta'y(t)$  becomes  $I(0)$ , where  $y(t)$  has to contain at least two  $I(1)$  variables. Under this assumption the identical VECM for (3.1) can be expressed as

$$A^*y(t-1) + A^{p-1}(L)\Delta y(t) = \varepsilon(t), \quad (3.16)$$

where  $y(t)$  contains both  $I(0)$  and  $I(1)$  variables,  $\Delta = (I - L)$ ,  $A^* = A^p(1)$ ,

$A^*y(t-1)$  is stationary, and

$$A^{p-1}(L) = I + \sum_{\tau=1}^{p-1} A^*_\tau L^\tau.$$

The first term in (3.16) [i.e.  $A^*y(t-1)$ ] is the error-correction term, which concerns the long-term cointegrating relationship. The second term in (3.16),  $A^{p-1}(L)\Delta y(t)$ , is referred to as the VAR part of the VECM, describing the short-term dynamics.

Because  $y(t)$  is cointegrated of order 1, the long-term impact matrix  $A^*$  must be singular. As a result  $A^* = \alpha\beta'$ , where  $\alpha$  and  $\beta$  are  $s \times r$  matrices and the rank of  $A^*$  is  $r$ ,  $r < s$ . The columns of  $\beta$  are the cointegrating vectors, and the rows of  $\alpha$  are the loading vectors.

Engle and Granger (1987) note that, for cointegrated systems, the VARs in first difference will be mis-specified and the VARs in levels will ignore important constraints on the coefficient matrices. Although these constraints may be satisfied asymptotically, efficiency gains and improvements in forecasts are likely to result by imposing them. Comparisons of forecasting performance of the VECMs versus VARs for cointegrated systems are reported in studies such as Engle and Yoo (1987) and LeSage (1990). The results of these studies consistently indicate that, in the short-term, there may be gains in using the unrestricted VAR models, but the VECMs produce long-term forecasts with smaller errors when the variables used in the models satisfy cointegration conditions. Subsequently Ahn and Reinsel (1990), and Johansen (1988, 1991) propose various algorithms for the estimation of cointegrating vectors in the full-order VECM models, which specify all non-zero entries in the coefficient matrices. Since the early 1990s, abundant literature has utilised the full-order

VECM models in analysing the short-term dynamics and the long-term cointegrating relationships for cointegrated time-series [see Reinsel and Ahn (1992), and Johansen (1992, 1995)].

VECM modelling has been increasingly employed to examine relationships in exchange rate and stock markets. For instance, in the context of foreign exchange markets, Granger *et al* (2000) conduct a bivariate VECM estimation to detect Granger causal relations between exchange rates and stock prices for five Asian countries. Diamandis *et al* (2000) examine the long-term properties of the monetary exchange rate model. They employ a VECM estimation using data for the Drachma-Dollar and Drachma-Mark exchange rates. Using equity markets, MacDonald and Power (1995) complete a VECM estimation in investigating cointegrating relations among stock prices, dividends and the retention term. Lee (1996) estimates a VECM using three I(1) cointegrated stock price, dividend and earnings series to explore the long-term relations among these financial variables. Barnhill *et al* (2000) employ a VECM to estimate cointegrating relations among investment grade indices, US Treasury yields, and default rates.

### 3.6.1 ZNZ Patterned VECM Modelling

As described in Section 1.3.1.1, in applications of VECM models to financial market data it may be assumed *a priori* that zero entries are required. In such cases the use of full-order VECM models may lead to incorrect inferences. Specifically, in conducting causality and cointegration analysis, if the entries assigned *a priori* to be zero were ignored and full-order VECM models were utilised, the power of statistical inferences would be weakened. Also, if the underlying true VECM and the associated cointegrating and loading vectors contained zero entries, the resultant specifications could produce different conclusions concerning the cointegrating relationships among the variables.

In addition, one difficulty encountered in empirical research using cointegration theory is to provide satisfactory financial and economic interpretation for estimated cointegrating vectors. As emphasised by Penm *et al* (1997) it is important to introduce *a priori* information, usually to produce ZNZ patterns. To explicitly address this issue Penm *et al* have presented a search algorithm in conjunction with model selection criteria to identify the optimal specification of a ZNZ patterned VECM for an  $I(1)$  system.<sup>13</sup> This VECM, with allowance for possible zero entries in the coefficient matrices, is referred to as a ZNZ patterned VECM. Given the optimal ZNZ patterned VECM, the number of cointegrating vectors can be confirmed. Once the ZNZ patterned impact matrix has been determined, along with the number of cointegrating vectors in the system, a tree-pruning procedure is then proposed to search for all acceptable ZNZ patterns of

---

<sup>13</sup> Of note, an  $I(1)$  system does not contain any fractionally integrated variables.

the cointegrating and loading vectors. After this, the dynamic ordinary least squares method suggested by Stock and Watson (1993) is utilised to estimate the acceptable patterned cointegrating vectors, and the regression method with linear restrictions as recommended in Penm *et al* (1997) is conducted to estimate the acceptable patterned loading vectors. Model selection criteria are again employed to determine the optimal ZNZ patterned cointegrating and loading vectors. This algorithm leads to a neat and effective analysis of the cointegrating relations in any vector time-series system and can be extended to higher order integrated systems [see Brailsford *et al* (2001c)].

This chapter indicates that full-order VECM models assume nonzero elements in all their coefficient matrices. As the number of elements to be estimated in these possibly over-parameterised models grows with the square of the number of variables, the degrees of freedom will be heavily reduced.

Application of VECM models to economic and financial time-series data have revealed that zero coefficient entries are indeed possible [see King *et al* (1991) and Penm *et al* (1997), (2001a)]. An optimal VECM specification with zero entries suggests that the cointegrating vectors and the loading vectors may also contain zero entries. However, the existence of zero entries has not been fully discussed in causality and cointegration theory. Specifically the ability to detect the presence or absence of indirect causality and/or Granger non-causality will be enhanced. Also the exact nature of the long-term cointegration relations will be crucially dependent upon finding those zero coefficient entries where the true structure does indeed include such zero entries.



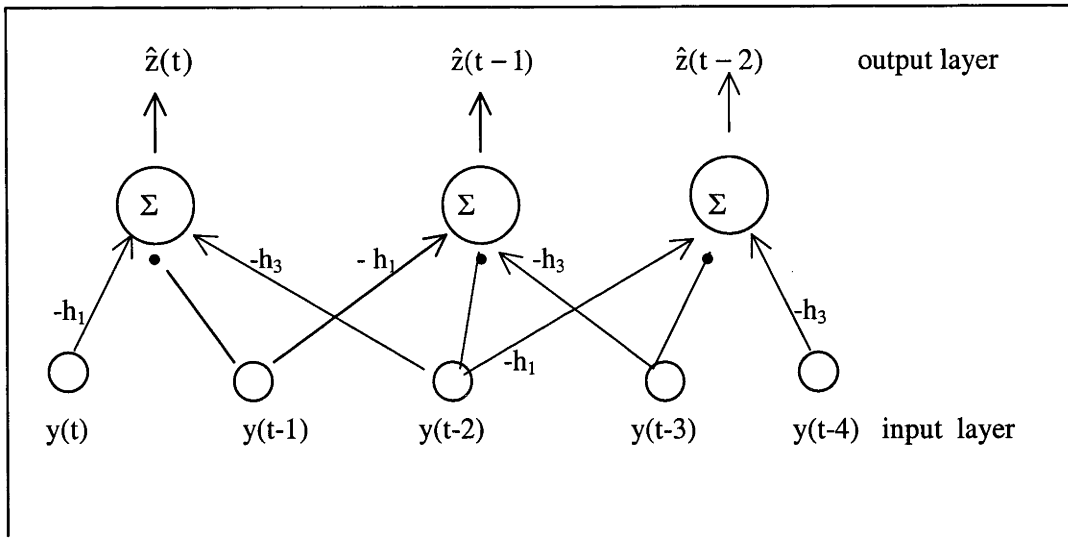
### 3.7 ZNZ Patterned Polynomial Neural Networks

A neural network is a mathematical model that processes information and generates some form of response based upon the relationship or pattern identified within the data. Neural networks also exist as computer-based systems containing many non-linear computational units or nodes interconnected by links with adjustable weights. Multi-layer networks with one or more hidden layers allow neural networks to classify functions that are not linearly separable. Historically neural networks were not taken seriously until they could solve non-linear problems. Several pruning algorithms for performance improvement have been proposed to eliminate non-significant connections. Can the new features of neural networks improve financial forecasting? This can only be addressed by conducting a detailed study on forecasting financial market prices.

Neural nets can be used as a tool for the study of time-series systems. A two-layered neural network proposed by Watanabe *et al* (1992) provides a typical example. It is simple, user-friendly and powerful. However, this net has difficulty in practical modelling of subset time-series systems. In this thesis the net is extended to increase its modelling power in the field of financial time-series analysis. Constraints on the connection strength (synaptic weights) are imposed on the network structure. Two types of connection (synapse), namely inhibitor arc and switchable connection, are incorporated into the neural net structure. Figure 3.1 shows both types of connection considered in the extended two-layered neural network for a specified illustrative subset VRDL model.

**Figure 3.1**

**An extended two-layered neural network<sup>a</sup>**



<sup>a</sup>This linear neural network computes the following 3rd order system:

$$\hat{z}(t) = -h_1 y(t) - h_3 y(t-2),$$

where  $r=1$ ,  $g=1$  and  $h_2 = 0$ . A switchable connection has an arrowhead and an inhibitor arc has a small solid black circle at the output node.

In Figure 3.1, the higher layer is the output layer, and the lower one is the input layer. All nodes in each layer express artificial neural units. Each unit in the output layer represents a neural output vector,  $\hat{z}(t)$  or its lagged vector. The  $\hat{z}(t)$  receives inputs from the input units represented by the current and the second lagged neural input-vectors. A switchable connection from an input node to the output node has an arrowhead at the output node, and the associated connection strength is switchable between zero and non-zero at any time. An inhibitor arc from an input node to the output node has a small solid black circle rather than an arrowhead at the output node. This circle means 'not connected' and the associated connection strength is constrained to zero at all times. Each

connection to  $\hat{z}(t)$  performs a linear transformation determined by the connection strength  $h_k$ , so that the total input for the output unit  $\hat{z}(t)$  is  $-h_1y(t) - h_2y(t-1) \cdots - h_p y(t+1-p)$ . If  $h_k = 0$  at  $t=T_1$ , the associated connection is inhibitive; if  $h_k \neq 0$  at  $t=T_1+dT$ , the associated connection is excitatory. It is noteworthy that  $dT$  is sufficiently large to ensure that the underlying relationships between  $z(t)$  and  $y(t)$  change smoothly and gradually. If, however,  $h_k = 0$  at all times, the associated connection becomes an inhibitor arc.

The extended network has a dynamic setting, that is, the 'presence and absence' restrictions on the coefficients of the optimal VRDL model may update each time a new observation becomes available. This 'presence and absence' pattern update indicates that some synapses interacting between neurons in the input layer and neurons in the output layer switch between the excitatory and inhibitive states. This specification is superior to the conventional static one in which no 'absence' restrictions are imposed on the coefficients and consequently all synapses are excitatory.

If the neural input-vector  $y(t)$  includes the first-order and second-order terms  $y_1(t)$ ,  $y_2(t)$ ,  $y_1(t)y_2(t)$ ,  $y_1^2(t)$  and  $y_2^2(t)$ , the VRDL model can be used to construct a three-layered polynomial neural network. The hidden-node transfer function in this network consists of a quadratic regression polynomial of two variables used by the group method of data handling (GMDH) algorithm of Ivakhnenko [see Farlow (1984)]. General connection between the mean-corrected input and output vectors can be expressed as

$$f_h(y_1(t), y_2(t)) = h_{11} y_1(t) + h_{12} y_2(t) + h_{13} y_1(t) y_2(t) + h_{14} y_1^2(t) + h_{15} y_2^2(t),$$

where  $y_i(t)$ ,  $i = 1, 2$ , are the input variables and  $h_{1i}$ ,  $i = 1, 2, \dots, 5$  are coefficients.

The proposed construction method is simple to use and can be applied to an  $M$ -layered polynomial neural network with hidden layer nodes in layer  $m \in [1, M-2]$ .

Figure 3.2 illustrates the structure of a polynomial network with a single hidden-layer for the predictor of a VRDL model:<sup>14</sup>

$$\hat{z}(t) = -h'_1 y(t) - h'_2 y(t-1) - h'_3 y(t-2), \quad (3.17)$$

where  $y(t) = [y_1(t) \ y_2(t) \ y_1(t)y_2(t) \ y_1^2(t) \ y_2^2(t)]'$  and  $h'_i = [h_{i,1} \ h_{i,2} \ \dots \ h_{i,5}]$ ,  $i=1, 2$  and  $3$ .

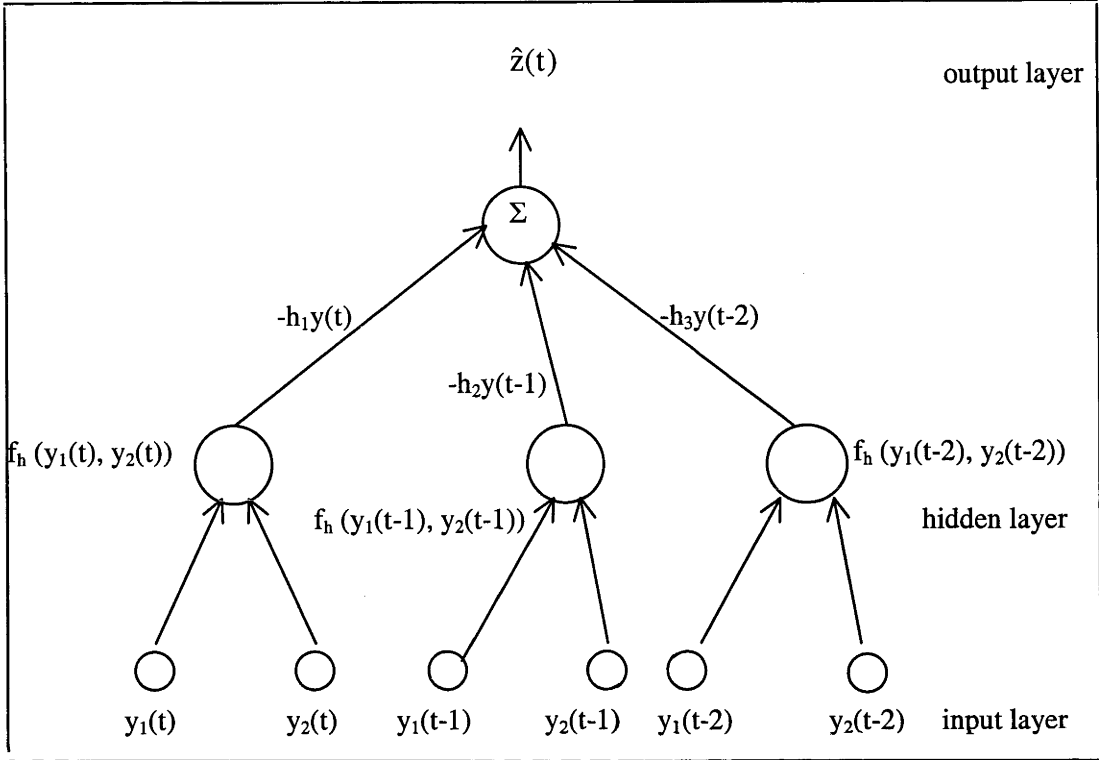
If  $y(t-1)$  in (3.17) is missing, (3.17) becomes a subset VRDL model. In this case, the hidden unit operating  $f_h(y_1(t-1), y_2(t-1))$  becomes inoperative and the corresponding incoming and outgoing arcs become inhibitive.

---

<sup>14</sup>Conventionally  $h_0, h_1, \dots$  denote the coefficients of  $y(t), y(t-1), \dots$  respectively. However in engineering literature, in order to enhance the presentation of recursive formulas,  $h_1, h_2, \dots$  are chosen to denote the coefficients of  $y(t), y(t-1), \dots$  respectively. The “engineering” approach is adopted when describing neural networks.

Figure 3.2

A polynomial neural network with a single hidden-layer for a VRDL model



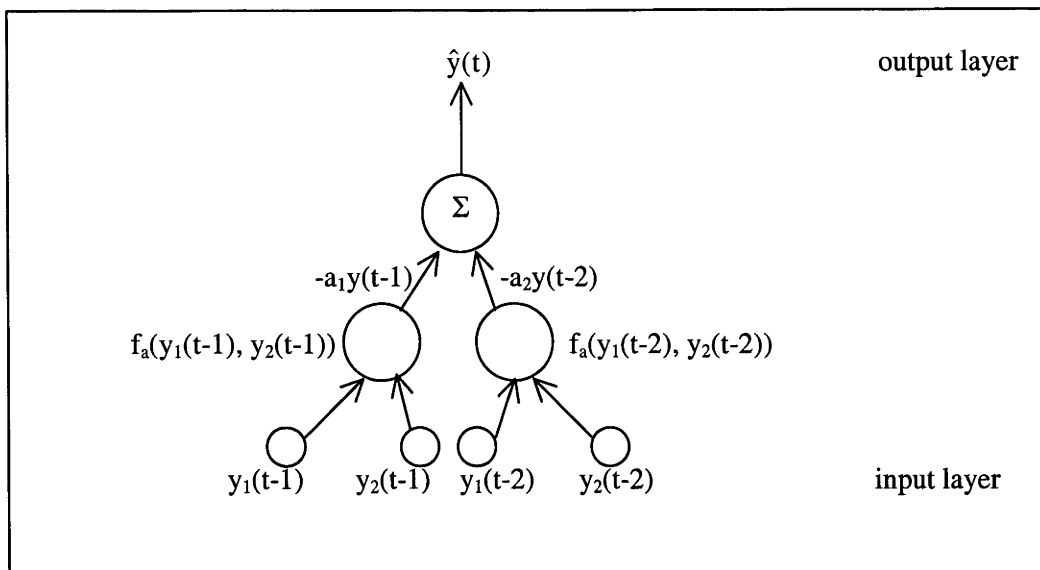
Analogously, Figure 3.3 illustrates the structure of a three-layered polynomial neural network for the predictor of a VAR model

$$\hat{y}(t) = -a_1 y(t-1) - a_2 y(t-2), \quad (3.18)$$

where  $a_i = [a_1(i) \ a_2(i) \ \dots \ a_5(i)]^T$ ,  $i=1$  and  $2$ .

**Figure 3.3**

**A three-layered polynomial neural network for a VAR model**



Neural nets with inhibitor arcs and switchable connections are intuitively the most direct approach to increasing the modelling power of neural nets. These extensions provide neural nets with an ability to model sequentially changing time-series systems with a subset structure. In Chapter 7 this thesis provides algorithms for both time and order updating which leads to the optimal synaptic weight updating and allows for the extended neural network which include the optimal dynamic node creation/deletion.

### 3.8 Summary and Linkage

Conventional full-order time-series models have become common in financial time-series analysis [Eun and Shim (1989), Bekaert *et al* (1999), Froot *et al* (1998), Granger *et al* (2000), Lee (1996), Barnhill *et al* (2000)]. However these models are typically constrained through problems of over-parameterisation.

In this chapter, a new ZNZ patterned time-series modelling approach comprising ZNZ patterned VAR modelling, ZNZ patterned VECM modelling and ZNZ patterned polynomial neural networks has been introduced. This approach is a more straightforward, certain and effective means of testing for causality and cointegrating relations, and conducting modelling and simulations. It can provide important insights into, and strengthen the modelling power of, financial time-series analysis. It also forms the basis of model building in the thesis. The modelling techniques developed are then applied to various tests and/or circumstances in exchange rate and equity markets.

The following chapter presents extensions and specific techniques of the models described in this chapter. It will demonstrate that the identified Granger causal relations among the variables from the use of the ZNZ patterned VAR models with unit roots are identical to the causal relations identified from the use of the equivalent ZNZ patterned VECM. It will also present an effective and efficient algorithm to select the optimal ZNZ patterned cointegrating and loading vectors in a ZNZ patterned VECM framework for an  $I(1)$  system. The algorithm can be

applied to a higher order integrated system. It is simple to use and leads to an efficient analysis of the cointegrating relationships in vector financial time-series.



## CHAPTER 4

### CAUSALITY DETECTION AND COINTEGRATION INVESTIGATION

#### 4.1 Introduction

In this chapter the necessary and sufficient condition to test for ‘overall causality’, that is the presence of Granger-causality, Granger non-causality and indirect causality in a ZNZ patterned VAR model, is discussed. It is argued in the next section that the coefficient patterns of the ZNZ patterned VAR model are a more straightforward and effective basis for detecting overall causality.

Causality detection in ZNZ patterned VECM models is also discussed in Section 4.3. It is shown that the identified Granger causal relations among the variables can be detected from the use of the ZNZ patterned VAR model with unit roots and from the use of the equivalent ZNZ patterned VECM.

Section 4.4 demonstrates an effective and efficient search algorithm to select from an  $I(1)$  system ZNZ patterned cointegrating and loading vectors in a ZNZ patterned VECM, when the long-term impact matrix contains zero entries. The algorithm can be applied to higher order integrated systems. Some concluding remarks are provided in Section 4.5.

## 4.2 Causality Patterns in ZNZ Patterned VAR Modelling

In VAR modelling of (3.1), consider a bivariate system where  $y(t) = [y_1(t) \ y_2(t)]'$ . Causal ordering can be defined using the work of Kang (1981).

Consider  $a_{ij}^p(L) = \sum_{\tau=1}^p \alpha_{ij}^{\tau} L^{\tau}$ , where  $a_{ij}^p(L)$  is the  $(i,j)$ -th entry of  $A^p(L)$ .

Definition (a):  $y_1(t)$  Granger non-causes  $y_2(t)$ , and  $y_2(t)$  Granger causes  $y_1(t)$  if and only if  $a_{21}^p(L) = 0$  and at least one  $\alpha_{12}^{\tau}$ ,  $\tau = 1, \dots, p$ , is nonzero.

That means

$$A^p(L) = \begin{bmatrix} a_{11}^p(L) & a_{12}^p(L) \\ 0 & a_{22}^p(L) \end{bmatrix}, \text{ and the coefficients, } \alpha_{12}^{\tau}, \tau = 1, \dots, p \text{ in } a_{12}^p(L) \text{ can}$$

be either zero or nonzero, but at least one  $\alpha_{12}^{\tau}$  is nonzero.

Further, there exist  $2^{p-1}$  different patterns of  $a_{12}^p(L)$  in this bivariate system, indicating that  $y_1(t)$  Granger non-causes  $y_2(t)$ , and  $y_2(t)$  Granger causes  $y_1(t)$ .

Definition (b):  $y_2(t)$  Granger non-causes  $y_1(t)$ , and  $y_1(t)$  Granger causes  $y_2(t)$  if and only if  $a_{12}^p(L) = 0$  and at least one  $\alpha_{21}^{\tau}$ ,  $\tau = 1, \dots, p$ , is nonzero.

That means

$A^p(L) = \begin{bmatrix} a_{11}^p(L) & 0 \\ a_{21}^p(L) & a_{22}^p(L) \end{bmatrix}$ , and the coefficients,  $\alpha_{21}^\tau$ ,  $\tau = 1, \dots, p$  in  $a_{21}^p(L)$  can

be either zero or nonzero, but at least one  $\alpha_{21}^\tau$  is nonzero.

Definition (c):  $y_2(t)$  Granger causes  $y_1(t)$ , and  $y_1(t)$  Granger causes  $y_2(t)$  if and only if  $a_{12}^p(L) \neq 0$  and  $a_{21}^p(L) \neq 0$ .

Definition (d):  $y_2(t)$  Granger non-causes  $y_1(t)$ , and  $y_1(t)$  Granger non-causes  $y_2(t)$  if and only if  $a_{12}^p(L) = 0$  and  $a_{21}^p(L) = 0$ .

The above causality patterns can be detected from the optimal selected ZNZ patterned VAR proposed in Penm and Terrell (1984a).

More general causal patterns can be treated using definitions suggested by Hsiao (1982). Consider the following trivariate system:

$$\begin{bmatrix} a_{11}^p(L) & a_{12}^p(L) & a_{13}^p(L) \\ a_{21}^p(L) & a_{22}^p(L) & a_{23}^p(L) \\ 0 & a_{32}^p(L) & a_{33}^p(L) \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = \varepsilon(t), \quad (4.1)$$

which describes  $y_1(t)$  causing  $y_3(t)$  but only through  $y_2(t)$ . In this trivariate system the above indirect causality implies  $a_{31}^p(L) = 0, a_{21}^p(L) \neq 0$  and  $a_{32}^p(L) \neq 0$ . Also  $\{\alpha_{12}^\tau = 0 \text{ or } \neq 0\}$ ,  $\{\alpha_{13}^\tau = 0 \text{ or } \neq 0\}$  and  $\{\alpha_{23}^\tau = 0 \text{ or } \neq 0\}$ ,  $\tau = 1, \dots, p$ . From (4.1), Hsiao (1982) indicates that the greater the number of

components,  $y_i(t)$ ,  $i = 1, 2, \dots$ , the more complicated are the causal patterns that may be detected.

#### **4.2.1 Empirical Causality Detection using VAR Modelling**

VAR modelling has been widely utilised in conducting causality detection tests. Caines *et al* (1981) propose procedures to test sales and advertising for Granger-causality in a class of VAR models. In an application to supermarket sales analysis, Caines *et al* use likelihood ratio tests to discriminate between various VAR models held as the null and alternative hypotheses respectively. A supermarket sales model is then constructed to aid supermarket managers in their decision-making.

Hsiao (1981) suggests a stepwise VAR modelling method of testing the supply of money and aggregate nominal income for Granger-causality. The variance-covariance matrices of various VAR models are estimated to calculate the log-likelihood values. Likelihood ratio tests are then carried out to select the optimal VAR model which is used as a basis for detecting Granger-causality.

Bar-Yosef *et al* (1987) use Hsiao's method to examine the linkage between corporate earnings and corporate investment. The various bi-variate AR models are constructed and their associated variance-covariance matrices are estimated to

test Granger-causality. The empirical results show that corporate earnings are a determinant of corporate investment.

In an investigation of measurement of Granger-causality, Geweke (1982) derives an interesting means of measuring the linear dependence and feedback present in multiple time series. Geweke also shows how the notions of causality relate to exogeneity in the context of a complete dynamic simultaneous equation model.

Bhattacharya *et al* (2000) utilise a sequential hypothesis testing procedure using standard likelihood ratio tests to analyse the causal direction of volatility transmission between two different classes of shares in the framework of a VAR. Lee (1992) uses a VAR approach to investigate causal relations and dynamic interactions among asset returns, real activity, and inflation in postwar USA. Bekaert and Hodrick (1992) examine the predictability of excess returns on equity and foreign exchange markets by using a six-variable VAR model. Whitelaw(1994) investigates a VAR for assessing the time series properties of the expected volatility of stock market returns. Thorbecke (1997) takes impulse-response functions from a VAR to analyse the relationship between monetary policy and stock returns. Bernanke and Blinder (1992) employ a VAR in which the federal funds rate is used to measure monetary policy.

### 4.3 Zero Entries in a ZNZ Patterned VAR and Its Equivalent VECM for an I(d) System

Full-order VAR models specify nonzero elements in all their coefficient matrices. As the number of elements to be estimated in these models grows with the square of the number of variables, the degrees of freedom will be heavily reduced. Moreover the statistical and numerical accuracy of coefficient estimates in these full-order models will be diminished, where the true structure does indeed include zero entries. As indicated in Penm and Terrell (1984a), the ZNZ patterned VAR model is a more straightforward and effective means of testing for Granger causal relations including Granger-causality, Granger non-causality and indirect causality.

However recent cointegration work [see Penm *et al* (1997)] suggests that, if cointegrating relations exist between the variables, then the use of the VECM model which is equivalent to a VAR model with unit roots, may be more effective for testing Granger-causality.

The equivalent VECM derived from (3.1) in an I(d) system can be presented as follows:

$$A^p(1)y(t-1) + A^{p-1}(1)\Delta y(t-1) + \dots + A^{p-d+1}(1)\Delta^{d-1}y(t-1) + A^{p-d}(L)\Delta^d y(t) = \varepsilon(t), \quad (4.2)$$

where  $A^{p-i}(1)\Delta^i y(t-1)$  are stationary,  $i=0, \dots, d-1$ . The first  $d$  terms are the error-correction terms, while  $A^{p-d}(L)\Delta^d y(t)$  is said to be the autoregressive part of the model.

Further, the following relations can be achieved:

$$A^k(L) = A^k(1)L + A^{k-1}(L)(I-L), k = p, p-1, \dots, p-d+1. \quad (4.3)$$

Since Granger-causality detection is crucially dependent on the positions of off-diagonal zero entries in the coefficient matrices, this discussion therefore focuses on the positions where  $i \neq j$ .

If the  $(i,j)$ -th entries of  $A^k(L)$ ,  $A^k(1)$ , and  $A^{k-1}(L)$  are  $a_{ij}(L)$ ,  $a_{ij}(1)$ , and  $c_{ij}(L)$  respectively, then the following relation is established:

$$a_{ij}(L) = a_{ij}(1)L + c_{ij}(L)(1-L), \quad i \neq j. \quad (4.4)$$

Now  $c_{ij}(L)$  is defined as a scalar polynomial with coefficients  $c_1, \dots, c_{k-1}$  by

$$c_{ij}(L) = c_1L + \dots + c_{k-1}L^{k-1},$$

and thus

$$c_{ij}(L)(1-L) = c_1L + (c_2 - c_1)L^2 + \dots + (c_{k-1} - c_{k-2})L^{k-1} - c_{k-1}L^k. \quad (4.5)$$

If  $a_{ij}(L) = 0$ , then  $a_{ij}(1)$  will also be zero. At (4.4) it was established that  $c_{ij}(L)(1-L) = 0$ , and (4.5) produces  $c_1 = 0$ ,  $c_2 - c_1 = 0$ ,  $\dots$ ,  $c_{k-1} - c_{k-2} = 0$ ,  $c_{k-1} = 0$ , which lead to  $c_i = 0$ ,  $i = 1, \dots, k-1$ , and therefore  $c_{ij}(L) = 0$ .

At this point, if the  $(i,j)$ -th entry of  $A^k(L)$  is zero, then the  $(i,j)$ -th elements of both  $A^k(1)$  and  $A^{k-1}(L)$  are zeros. Therefore, if it is shown that if every  $(i,j)$ -th entry is zero for all coefficient matrices in a VAR then all  $(i,j)$ -th coefficient elements in the error-correction terms and in the vector autoregressive part of the VECM, will also be zeros.

Analogously it is evident that if the  $(i,j)$ -th elements of all  $A^k(1)$ ,  $k = p, p-1, \dots, p-d+1$  and  $A^{k-1}(L)$  in (4.2) are zeros then the  $(i,j)$ -th entry of  $A^p(L)$  in the equivalent VAR will be zero. Therefore this thesis can conclude that if all  $(i,j)$ -th coefficient elements in the error-correction terms and all  $(i,j)$ -th coefficient elements in the vector autoregressive part of the VECM are zeros, then every  $(i,j)$ -th entry is zero for all coefficient matrices in a VAR.

The implications of the above proof are obvious. If  $y_j$  does not Granger-cause  $y_i$  then every  $(i,j)$ -th entry must be zero for all coefficient matrices in the VAR. Also all  $(i,j)$ -th coefficient elements in the equivalent VECM are zeros.

Further, (4.3) can be expressed as follows:

$$A^p(L) = A^p(1) + A^{p-1}(L) - A^{p-1}(L)L \quad (4.6.1)$$

$$A^{p-1}(L) = A^{p-1}(1) + A^{p-2}(L) - A^{p-2}(L)L \quad (4.6.2)$$

$\vdots$

$$A^{p-d+1}(L) = A^{p-d+1}(1) + A^{p-d}(L) - A^{p-d}(L)L. \quad (4.6.3)$$



From (4.6.3) it is obvious that if the  $(i,j)$ -th element of  $A^{p-d+1}(1)$  is nonzero, then the  $(i,j)$ -th element of  $A^{p-d+1}(L)$  is nonzero. Also if the  $(i,j)$ -th element of  $A^{p-d}(L)$  is nonzero, then a zero  $(i,j)$ -th element of  $A^{p-d+1}(1)$  leads to a nonzero  $(i,j)$  element of  $A^{p-d+1}(L)$ . Thus, it has been proved that if there exists a nonzero  $(i,j)$ -th element in either  $A^k(1)$  or  $A^{k-1}(L)$ ,  $k=p, p-1, \dots, p-d+1$  in (4.3), then the  $(i,j)$ -th element of  $A^k(L)$  is nonzero. This outcome shows that if any single  $(i,j)$ -th element is nonzero in any one of the  $d$  matrices,  $A^k(1)$ ,  $k= p, p-1, \dots, p-d+1$ , or  $A^{p-d}(L)$  in the VECM in (4.2) is nonzero, then the  $(i,j)$ -th element of  $A^p(L)$  in the equivalent VAR is nonzero.

Analogously from (4.6.1) if the  $(i,j)$ -th element of  $A^p(L)$  is nonzero, then at least the  $(i,j)$ -th element is nonzero in one of the following  $d$  coefficient matrices, or  $A^{p-d}(L)$ :

$$A^p(1), A^{p-1}(1), \dots, A^{p-d+1}(1).$$

Therefore if  $y_j$  does Granger-cause  $y_i$ , then the  $(i,j)$ -th element of  $A^p(L)$  in the VAR is nonzero. In addition at least a single  $(i,j)$ -th coefficient element is nonzero in  $A^p(1), A^{p-1}(1), \dots, A^{p-d+1}(1)$ , or  $A^{p-d}(L)$  in the equivalent VECM.

An indirect causality from  $y_j$  to  $y_i$  through  $y_m$  indicates  $y_j$  causing  $y_i$  but only through  $y_m$ . Hence,  $y_j$  Granger-causes  $y_m$ ,  $y_m$  Granger-causes  $y_i$ , and  $y_j$  does not Granger-cause  $y_i$  directly. It can easily be demonstrated that the VAR in (3.1) has nonzero  $(m,j)$ -th and  $(i,m)$ -th elements and a zero  $(i,j)$ -th element of  $A^p(L)$ . The equivalent indirect causality can also be shown in the equivalent VECM.

It is noteworthy that Johansen (1988) proposed the following VECM equivalent to the VAR model of (3.1) in an I(1) system:

$$H^{p-1}(L)\Delta y(t) + A^* y(t-p) = \varepsilon(t), \quad (4.7)$$

where  $H^{p-1}(L) = I + \sum_{i=1}^{p-1} H_i L^i$ .

The error-correction term of this VECM is  $A^* y(t-p)$ , while the error-correction term in (3.16) is  $A^* y(t-1)$ .

Thus

$$H_k = I_p + A_1 + \dots + A_k, \quad k = 1, \dots, p-1, \quad (4.8)$$

$$\text{and } A^* = I_p + A_1 + \dots + A_p. \quad (4.9)$$

It is recalled that  $a_{ij}^k$  denotes the  $(i,j)$ -th entry of  $A_k$ ,  $g_{ij}^k$  and  $a_{ij}^*$  can denote the  $(i, j)$ -th entry of  $H_k$  and  $A^*$  respectively. From (4.8) it is obvious that all  $p$  entries,  $\{a_{ij}^k, k = 1, \dots, p\}$  are zeros; then  $a_{ij}^*$  is zero and all  $\{g_{ij}^k, k = 1, \dots, p-1\}$  are also zeros. Similarly if  $a_{ij}^*$  and all  $\{g_{ij}^k, k = 1, \dots, p-1\}$  are zeros, then  $\{a_{ij}^k, k = 1, \dots, p\}$  are zeros.

Therefore if  $y_j$  does not Granger-cause  $y_i$ , then every  $(i,j)$ -th entry is zero for all coefficient matrices in the VAR. Also all  $(i,j)$ -th coefficient elements in the equivalent VECM are zeros.

If any single  $a_{ij}^k, k = 1, \dots, p$  in the VAR is nonzero, then the  $(i,j)$ -entry is nonzero in either  $A^*$  or  $H^{p-1}(L)$ .

Expression (4.8) can be rewritten as

$$\begin{aligned} H_1 &= I_p + A_1, \\ H_2 &= H_1 + A_2, \\ &\vdots \\ H_k &= H_{k-1} + A_k, \quad k = 2, \dots, p-1. \end{aligned} \tag{4.10}$$

From (4.10) if  $a_{ij}^1$  is nonzero, then  $g_{ij}^1$  is nonzero. However if  $a_{ij}^1$  is zero, then  $g_{ij}^1$  will be zero. Similarly, if  $a_{ij}^2$  is nonzero, then  $g_{ij}^2$  is nonzero. In addition, it is obvious that if  $a_{ij}^p$  is nonzero and all  $a_{ij}^k, k = 1, \dots, p-1$  are zeros, then  $a_{ij}^*$  is nonzero, although all  $g_{ij}^k, k = 1, \dots, p-1$  are zeroes.

Analogously it can be proved that if any single  $(i,j)$ -entry is nonzero in either  $A^*$  or  $H^{p-1}(L)$ , then the  $(i,j)$ -th entry of  $A^p(L)$  in the equivalent VAR is nonzero.

Therefore if  $y_j$  does Granger-cause  $y_i$  in (4.7), then the  $(i,j)$ -th element of  $A^p(L)$  in the VAR is nonzero. In addition the  $(i,j)$ -entry is also nonzero in either  $A^*$  or  $H^{p-1}(L)$  in the equivalent VECM.

#### 4.4 Selection of ZNZ Patterned Cointegrating Vectors and Loading Vectors in ZNZ Patterned VECM Modelling

As noted in Section 3.3, the VAR model is described as:

$$y(t) + \sum_{\tau=1}^q A_{\tau} y(t - \tau) = \varepsilon(t), \quad (4.11)$$

where  $\varepsilon(t)$  is a  $s \times 1$   $I(0)$  vector process with  $E\{\varepsilon(t)\} = 0$  and

$$\begin{aligned} E\{\varepsilon(t)\varepsilon'(t - \tau)\} &= G, & \tau &= 0, \\ &0, & \tau &> 0, \end{aligned}$$

$A_{\tau}$ ,  $\tau = 1, 2, \dots, q$  are  $s \times s$  parameter matrices, and

$$A^q(L) = I + \sum_{\tau=1}^q A_{\tau} L^{\tau}.$$

$L$  denotes the lag operator and the roots of  $|A^q(L)| = 0$  lie outside or on the unit circle.

Further,  $y(t)$  is said to be  $I(1)$ , if it contains at least one element which must be differenced before it becomes  $I(0)$ . Then  $y(t)$  is said to be cointegrated of order 1 with the cointegrating vector,  $\beta$ , if  $\beta'y(t)$  becomes  $I(0)$ , where  $y(t)$  has to contain at least two  $I(1)$  variables. Under this assumption the associated VECM for (4.11) can be expressed as follows:

$$A^*y(t-1) + A^{q-1}(L)\Delta y(t) = \varepsilon(t), \quad (4.12)$$

where  $y(t)$  contains both  $I(0)$  and  $I(1)$  variables,  $\Delta = (I-L)$ ,  $A^* = A^q(1)$ ,  $A^q(1)y(t-1)$  is stationary, and

$$A^{q-1}(L) = I + \sum_{\tau=1}^{q-1} A_{\tau}^* L^{\tau}.$$

The first term in (4.12) is the error-correction term.  $A^{q-1}(L)\Delta y(t)$  is referred as the VAR part of the VECM.

Because  $y(t)$  is cointegrated of order 1, the long-term impact matrix,  $A^*$ , must be singular. As a result,  $A^* = \alpha\beta'$ , where  $\alpha$  and  $\beta$  are  $s \times r$  matrices and the rank of  $A^*$  is  $r$ . The columns of  $\beta$  are the cointegrating vectors, and the rows of  $\alpha$  are the loading vectors.

One problem encountered in empirical research using cointegration theory is to provide satisfactory financial and economic interpretation of estimated cointegrating vectors. As demonstrated by Wickens (1996) this is often difficult without introducing *a priori* information particularly where this *a priori* information determines the presence or absence of certain coefficients. To explicitly address this issue Section 3.6.1 presents a search algorithm to identify the specification of a VECM with ZNZ patterned cointegrating and loading vectors. This algorithm allows zero coefficients in the VECM including the cointegrating vectors, the loading vectors and the VAR part of the VECM. The

specification so determined provides a useful basis for financial and economic interpretation of the long-term equilibrium relationships as well as the short-term dynamics.

#### 4.4.1 Search Algorithm

In the proposed algorithm for an  $I(1)$  system, the identification of ZNZ patterned  $A^*$  and the determination of ZNZ patterned  $\alpha$  and  $\beta$  are carried out in the following way. First, model selection criteria are used to select the optimal subset VECM with zero entries to determine the ZNZ patterned  $A^*$ .

Second, after the ZNZ patterned  $A^*$  is determined, the rank of the matrix  $A^*$  is then computed using the singular value decomposition (SVD) method, and the number of cointegrating vectors in the system will be known.

Third, given that the ZNZ patterned  $A^*$  has been determined and the rank of  $A^*$  has been computed, it is then possible to proceed with the tree-pruning algorithm as adapted for an  $I(1)$  system to obtain all acceptable ZNZ patterned  $\alpha$ s and  $\beta$ s which are consistent with the ZNZ patterned  $A^*$ . Let  $\alpha_p$  and  $\beta_p$  denote a ZNZ pattern of  $\alpha$  and  $\beta$  respectively and  $A_p$  the ZNZ pattern of  $A^*$ . If the  $(i,j)$ -th entry of the product,  $\alpha_p \beta_p'$  is zero, and the corresponding  $(i,j)$ -th entry of  $A_p$  is also zero, then both  $\alpha_p$  and  $\beta_p$  are acceptable. This tree-pruning algorithm, which avoids the need to evaluate all possible ZNZ patterned  $\alpha$ s and  $\beta$ s, is discussed in Appendix B.

The ZNZ patterns of acceptable  $\alpha$ s and  $\beta$ s depend on the pattern of  $A^*$  determined earlier by model selection criteria. Of note, the imposition of zero entries on  $\beta$  does not preclude a similar restriction on  $\alpha$ . One example is that if the determined  $A^*$  contains a zero row, such as:

$$A^* = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix},$$

where 1 denotes a non-zero entry. In this case zero restrictions will have to be imposed on the first row of  $\alpha$ . This is because the pattern of  $A^*$  implies that the cointegrating relations in the system have no influence on the first variable in the system. Noting that the number of zeros in  $\alpha$  and  $\beta$  are not fixed even with a given ZNZ patterned  $A^*$ , many differently patterned  $\alpha$ s and  $\beta$ s can be obtained using the tree-pruning algorithm. A simple example can be used for demonstration.

$$\text{Let } A^* = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix},$$

where the rank of  $A^*$  is 2.

At least three candidate sets of  $\alpha$  and  $\beta$  can be obtained, which are:

$$\alpha = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{I})$$

$$\alpha = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{II})$$

$$\alpha = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{III})$$

The cointegrating relationships implied by (I), (II) and (III) are different. While (I) and (II) imply that  $y_1$ ,  $y_2$  and  $y_3$  are cointegrated, (III) indicates that  $y_1$  and  $y_2$  are cointegrated and  $y_3$  is an  $I(0)$  series. It is obvious that this thesis cannot take the zero-maximising approach of choosing the  $\beta$  with the maximum number of zero entries to determine the ZNZ patterns of  $\alpha$  and  $\beta$ . If it did, then (III) would be selected, not (I) nor (II), while the true model could be either (I) or (II). As a result it again utilises model selection criteria to assess (I), (II) and (III), and then select the optimal ZNZ patterns for  $\alpha$  and  $\beta$ . Although (I) and (II), in theory, both indicate that  $y_1$ ,  $y_2$  and  $y_3$  are cointegrated, in practice different forecasting performance will result from (I) and (II). Using model selection criteria in this situation will aid in the selection between (I) and (II) in terms of forecasting performance.

To obtain the correct specification for  $\alpha$  and  $\beta$ , it is necessary to determine whether  $\alpha$  and  $\beta$  can be uniquely obtained by factorising  $A^*$ . If it is possible, the factorisation can be carried out. If not, the efficient estimation of  $I(1)$  cointegrated systems based on a triangular ECM representation [see Stock and Watson (1993)] can be employed to estimate  $\beta$ . Since any non-zero entry in  $\beta$  could be normalised as unity, it can repeat the estimation procedure with all possible normalisations. The normalisation, which produces the smallest value for the model selection criteria, is then selected as the candidate  $\beta$ . After the optimal normalisation is determined for every candidate  $\beta$ , it is possible to estimate the associated acceptable ZNZ patterned



$\alpha$ s. Consequently the optimal  $\alpha$  and  $\beta$  are the ones which result in the minimum value for model selection in the VECM framework.

#### **4.5 Summary and Linkage**

This chapter has established that both conventional full-order VAR and VECM models contain all non-zero entries in their coefficient matrices, As the number of parameters to be estimated in these full-order models grows with the square of the number of variables, the degrees of freedom will be heavily reduced. Moreover the statistical and numerical accuracy of the parameters estimated will be diminished, where the true structure does indeed include zero parameters. Therefore the modelling power of the full-order approach for causality detection and cointegration investigation is weakened.

This chapter has also demonstrated that the optimal ZNZ patterned VAR model can be used as a basis for detecting causality for stationary vector financial time-series. However if cointegrating relations exist between the variables, then the use of the equivalent patterned VECM model may be more effective for testing Granger-causality. An effective and efficient algorithm is also shown to select the optimal patterned VECM for an I(1) system.

The following chapter presents applications of the ZNZ patterned VAR modelling described in this chapter. Two issues are investigated concerning causality detection, using the modelling techniques developed in this chapter. The first

issue analyses the dominant factors influencing the Euro's exchange rate movements through the development of ZNZ patterned VAR models. The second study utilises ZNZ VAR modelling of the Hong Kong stock market, including the impact of the Euro on the market.

## **CHAPTER 5**

### **CAUSAL ANALYSIS OF THE MONEY SUPPLY AND THE EURO AND ITS IMPACT ON THE HONG KONG STOCK MARKET**

#### **5.1 Introduction**

With the introduction on 1 January 1999 of the single European currency, the Euro became the official currency in the eleven participating countries of the European Union (EU). The increasing popularity of the Euro as a pegging currency reflects the internationalisation of the Euro. Also the Euro has been used as the largest weighting element in a basket of currencies for foreign exchange arrangements adopted by several Central European countries. The Euro has become the second most widely traded currency at the international level, behind the US Dollar and ahead of the Japanese Yen. In this chapter the ZNZ patterned VAR modelling is utilised to investigate direct Granger causal relations between the money supply of the Euro area, which comprises all participating countries of the EU and the Euro exchange rate. In addition, the hypothesis that the Euro exchange rate is a major influence on international stock markets is tested. The hypothesis is tested by examining the cause and effect relationship between the Hong Kong stock market and foreign exchange markets using the ZNZ patterned VAR modelling.

The remainder of this chapter is organised as follows. Section 5.2 reviews the background information on money supply in the Euro area and the Euro's potential influence on stock markets. Section 5.3 describes the data. Section 5.4 investigates the causal relationships between the movements in the Euro's exchange rate and the money supply. Section 5.5 examines the Euro's impact on the Hong Kong stock market. A summary is provided in Section 5.6.

## **5.2 Money Supply in the Euro Area and the Euro's Implications for the Hong Kong Stock Market**

### **5.2.1 Introduction to the Euro**

The introduction of the Euro has been a significant recent event in global financial markets. The Euro is intended to create broader, deeper and more liquid financial markets in Europe, and thus its main purpose is to improve the price stability and productivity of the European economy. Rather than experiencing constant fluctuations in the member exchange rates there will be a more consistent and predictable environment for international trade. Another reason why the European Central Bank introduced the Euro is based on its belief that the new currency will foster low inflation.

The Euro has already established itself as a credible and important currency in the world. To date the Euro/Dollar trading has been very active in the world's foreign

exchange markets through a wide range of instruments, offering significant hedging possibilities.

Over the period January 1999 to December 2000 the relative weakness of the Euro was a significant feature in international foreign exchange markets. During this period the value of the Euro relative to the US Dollar, in general, fell. The Euro's weakness throughout this period confounded earlier general expectations that it would trend upwards relative to the US Dollar [see ECB (2001a)].

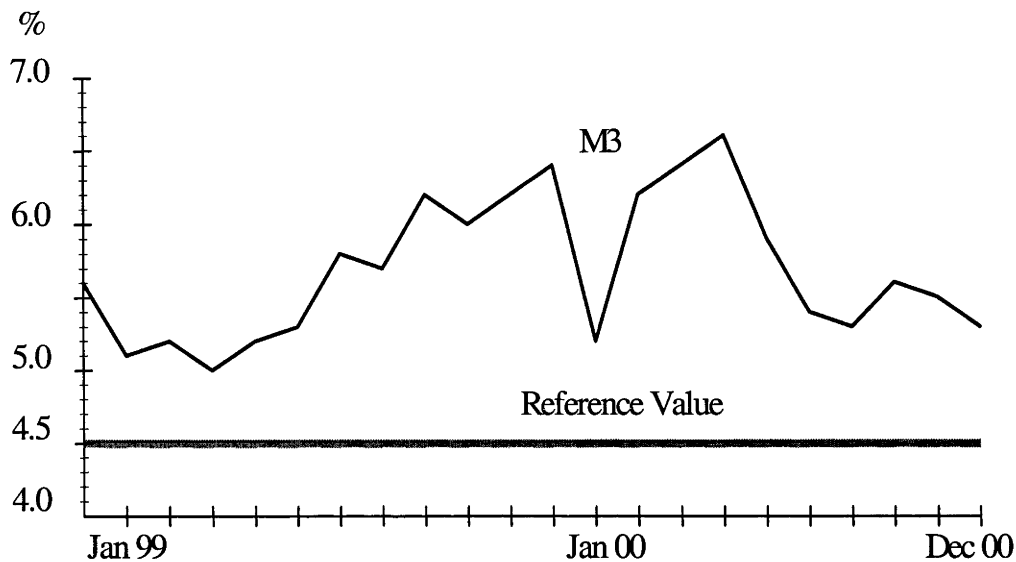
### **5.2.2 European Money Supply**

Money supply in the Euro area is measured by the standard stock of money (M3). It consists of short-term deposits, shorter deposits of up to 2 years, and marketable instruments. Figure 5.1 shows that over the period 1999 to 2000 the monthly measures of M3 have always been higher than the reference value of 4.5 percent set by the Governing Council of the European Central Bank (source: DataStream™). That is the growth rate of M3 has exceeded the 4.5 percent benchmark for the entire period.<sup>15</sup> The Governing Council adopts a price stability-oriented monetary policy strategy for the Eurosystem. That is the rate of monetary expansion is set to achieve the objective of price stability.

---

<sup>15</sup> The growth rate of M3 is measured relative to the previous month. It fell back in January 2000. This fall may have been caused by the general rises in interest rates during 1999 [ECB (2000)].

**Figure 5.1**  
**Growth of M3 in the Euro area**



Source: ECB

Money supply is linked to forex price movements. In the gold standard era, since the gold reserves of a country were limited, the growth rate of money supply was managed with close attention to the country's reserves. Unmanaged growth of money supply would lead to a depreciation of the country's currency.

In a floating exchange rate system, currencies fluctuate according to supply and demand. One tool that has been used to manage the exchange rate is through the money supply. However, such action can only be successful in the short-term. Governments are not able to control the exchange rate over a long period without regard to economic fundamentals.

The most widely held view is that, *ceteris paribus*, an expansion in money supply leads to a decrease in domestic interest rates. For a given expected inflation rate

this leads to a depreciation in the domestic currency. In the overshooting hypothesis, the immediate depreciation of the spot exchange rate will temporarily exceed, or overshoot, that of the long-term equilibrium exchange rate. Conversely, a tightening of monetary policy can lead to an appreciation of the domestic currency.

Lewis (1993) utilises VAR modelling to investigate the impact of US monetary shocks on the US Dollar exchange rate. The findings indicate that a loosening of monetary policy is associated with a depreciating currency. Evans (1994) assesses the impact of monetary shocks on exchange rate movements in the US, Germany and Japan. His conclusions are that a shock to the Federal fund rate has a stronger effect on the exchange rate than a shock to the interest differential. Eichenbaum and Evans (1995) investigate the effects of money shocks on the US dollar exchange rate. Their results show that monetary policy is important in explaining exchange rate movements, but they do not explain the majority of these movements. Cushman and Zha (1997) examine the effects of monetary shocks on the Canadian Dollar exchange rate movements. They conclude that a contraction in US monetary supply leads to an appreciation of the US Dollar against the Canadian Dollar. The four above-mentioned studies do not give any evidence to support the overshooting hypothesis. However the findings of Bonser-Neal *et al* (1998) support the overshooting hypothesis. They use event study methodology to investigate the impact of monetary shocks on exchange rates. Their findings suggest that the immediate response of the exchange rate to US monetary policy is

statistically and economically significant in most cases, and the overshooting hypothesis is acceptable in seven of the eight cases they examine [Bonser-Neal *et al* (1998)].

There is already literature examining the causal relationships between the money supply and economic activity in the Euro area [see BIS (2000)]. However an investigation into the direct causal relationships between the money supply and the Euro exchange rate has so far not been attempted. Thus the first major area of interest in this chapter is to investigate whether there is a direct causal relationship between the movements of the Euro's exchange rate and the money supply.

### **5.2.3 Forex and Stock Markets**

This section examines the linkage between the forex and stock markets. It seeks to explain why the Euro could impact on world stock markets. It uses Hong Kong as an example to examine the Euro's influence.

First, previous evidence has shown that exchange rate changes have a significant impact on stock prices, implying the former do contain relevant information about stock prices. For instance, Froot *et al* (1998) demonstrate that flows of capital influence exchange rate movements and such flows have been shown to be related to equity returns.<sup>16</sup> The and Shanmugaratnam(1992) and Yip (1996) have studied

---

<sup>16</sup> Also the relationship between exchange rates and stock prices is more complex than implied here and involves consideration of parity conditions and inflationary expectations. Nevertheless while exchange rate risk should not be separately priced if purchasing power parity holds, in the short-to-medium term, deviations from PPP have been reported [see Adler and Lehman (1983); Frenkel (1981)]. Under these conditions



the Singapore stock market and find that, in the small and open economy of Singapore, a strong Singapore Dollar is related to positive returns in the Singapore stock market. They emphasise that a strong Singapore Dollar lowers input costs, and thus limits imported inflation. Consequently productivity of Singapore based firms is improved, which is associated with a rise in the stock prices. Of note, Hong Kong and Singapore share many similar economic features in East Asia. Therefore we hypothesise a positive relationship between the Hong Kong Dollar and the Hong Kong stock market.

Second, as described in Section 5.2.1, during the test period 1 January 1999 to 31 December 1999 the weakness of the Euro has been a significant feature in international foreign exchange markets. The Annual Report of BIS in (2000) observed: "The Eurosystem had indicated that it would not react automatically to deviations of money growth from the reference value".

During the test period, interest rates on the marginal lending facilities were maintained by the ECB at 3.5 percent [see ECB (1999b)], while the interest rate in the US was about 6 percent during the same period, thereby creating a flow of capital from the Euro area to other markets, including Hong Kong. During this period the best lending rate in Hong Kong was above 8 percent [see Hong Kong Monetary Authority (2001)].

---

deviations from purchasing power parity will be priced to the extent that they represent exchange rate risk that must be borne by investors [see Jorion (1991); Dumas and Solnik (1995)]. In any event the purpose here is to illustrate how this analysis can provide insights into Granger-causal relationships among financial variables.

Third, Hong Kong is an important Asian banking and financial centre and generally allows free entry and exit of international funds. The Hong Kong stock market is the second largest in Asia after Tokyo. Also the Hong Kong government was the largest player in the Hong Kong stock market in 1999.<sup>17</sup> Further international investors are willing to place their money in Hong Kong stocks while assessing other investment opportunities, in particular those of investing in China. Hence it is hypothesised that the decline in the Euro had implications for Hong Kong's stock market in 1999.

The Hong Kong Monetary Authority used, as it still does, the linked exchange rate of HK7.8 Dollars to one US Dollar to encourage stability and investor confidence during and after the unification of Hong Kong with China. However in Hong Kong only the Hong Kong Monetary Authority uses this linked exchange rate, so this rate does not apply to other dealers in Hong Kong, who are subject to fluctuations in the Hong Kong and US Dollar exchange rate.

## 5.2 Data

To investigate the causal relationships between movements in the Euro's exchange rate and the money supply, monthly average data on the Euro's exchange rate ( $E_e$ ) and seasonally adjusted M3 are collected from DataStream<sup>TM</sup>

---

<sup>17</sup> During this year the Hong Kong government set up a fund to purchase shares valued at 118 billion Hong Kong Dollars from the local stock market [see Hong Kong Monetary Authority (2000)].

over the period September 1996 to December 2000. To examine stationarity for each series Microfit 4.0 is used to carry out the augmented Dickey-Fuller (ADF) unit root test. The results indicate that both  $\log E_e$  and  $\log M3$  are non-stationary.

To examine the Euro's impact on the Hong Kong stock market, all data are sampled daily between 1 January and 31 December 1999 from DataStream™. The Hang Seng Index (HSI) is used to proxy the Hong Kong stock market. It is the main stock market indicator in Hong Kong. This index comprises 33 constituent stocks which are the largest in the market. The aggregate market capitalisation of these stocks accounts for about 70 percent of the total market capitalisation on Hong Kong's stock exchange. At the beginning of 1999 the HSI was 9,000. However it climbed to about 17,000 by the end of 1999, closing with an 89 percent gain over the year.

Within this context, the following three variables are studied contemporaneously in a stochastic vector system using the ZNZ patterned vector AR modelling proposed above:

- (i) Euro to US Dollar - exchange rate (EUFX)
- (ii) Hong Kong's Hang Seng - stock price index (HSI)
- (iii) Hong Kong Dollar to US Dollar - exchange rate (HKFX).

Graphs of EUFX, HSI and HKFX are shown in Figures 5.2 to 5.4 respectively. The variables are log transformed such that  $y_1(t)=\log(\text{EUFX})$ ,  $y_2(t)=\log(\text{HSI})$  and  $y_3(t)=\log(\text{HKFX})$ . Following Penm and Terrell (1984a), Forsythe's (1957)

method is initially used for generating orthogonal polynomials to assess the data for suitable detrending to produce stationarity. The results show that detrending using a first-order polynomial is required before fitting the VAR models. The standard errors of estimates of coefficients are reported in Table 5.1. Thus all three variables are mean-corrected and detrended to achieve stationarity.

**Figure 5.2**

**Euro to US Dollar - exchange rate (EUFX), daily: 1 January 1999 to 31 December 1999**



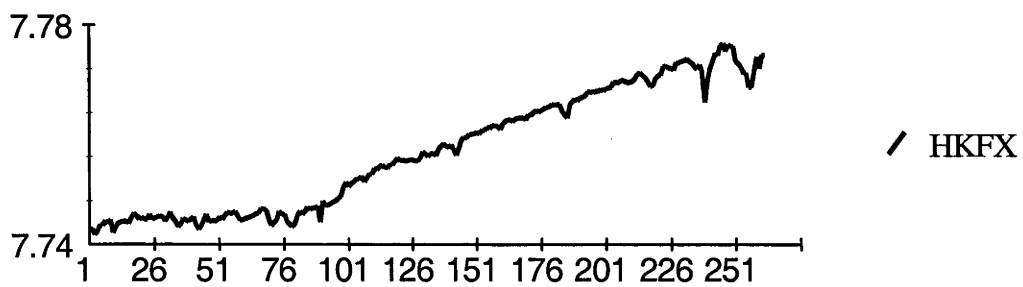
**Figure 5.3**

**Hang Seng stock price index (HSI), daily: 1 January 1999 to 31 December 1999**



**Figure 5.4**

**Hong Kong Dollar to US Dollar - exchange rate (HKFX), daily: 1 January 1999 to 31 December 1999**



**Table 5.1**  
**Orthogonal polynomial regression<sup>a</sup>**

	Intercept	Orthogonal polynomial P <sub>1</sub>
log(EUFX)	-0.11 (2.83E-03)	3.98E-04 (1.87E-05)
log(HSI)	9.20 (9.32E-03)	1.75E-03 (6.14E-05)
log(HSFX)	2.04 (3.17E-05)	1.50E-05 (2.09E-07)

<sup>a</sup>The values in parentheses are standard errors of the coefficient estimates

#### **5.4 The Causal Relationship between the Movements of the Euro's Exchange Rate and the Money Supply**

In detecting the causal relationships between the movements of the Euro's exchange rate and the money supply, the identification algorithms for ZNZ patterned VAR modeling as proposed in Chapter 3 are utilised to select the optimal VAR models for both mean-corrected and detrended  $\log(M3)$  and  $\log(E_e)$  at  $T=48, 49, 50, 51$  and  $52$ . The five cases correspond to August, September, October, November and December 2000 respectively. To demonstrate the usefulness of the proposed algorithms in a small sample environment, a maximum order of 12 is selected to cope with this small sample environment. Following the proposed algorithms, the optimal ZNZ patterned VAR models from  $T=48$  to  $T=52$

are chosen by using both the Hannan-Quinn Criterion (HQC) and the Schwarz Criterion (SC). The optimal models selected are estimated using the GLS techniques and are shown in Table 5.2. These models are then used as the benchmark models for analysing the causal relationships.

The patterned VAR selected by both criteria at all times shows Granger-causality from M3 to  $E_e$ , and Granger no-causality from  $E_e$  to M3.<sup>18</sup> This outcome confirms that M3 is an independent source of financial and economic disturbance and is influential over movements of the Euro's exchange rate during the test period. A change in M3 causes changes in the value of the Euro. These results are consistent with both theory and prior evidence.

---

<sup>18</sup> For comparison purposes VECM modelling to both  $\log E_e$  and  $\log M3$  without detrending is also conducted. The outcome indicates the selected optimal VECM comprises a lag one term in the autoregressive part. The lag coefficient matrix has a nonzero (1,2)-entry and a zero (2,1)-entry. Thus the one-way causality from M3 to  $E_e$  at  $T = 48, 49, 50, 51$  and  $52$  is confirmed.

**Table 5.2**

**The VARs <sup>a, b</sup> selected by both HQC and SC for detecting the causal relationships between E<sub>e</sub> and M3**

Sample size (T)	Non-zero lag coefficient structure for $y(t) = [\log E_e, \log M3]'$	Pattern of Granger causality <sup>c</sup>
48	lag 1 $\begin{bmatrix} -0.825 & 0.178 \\ (0.073) & (0.090) \\ 0 & -0.806 \\ & (0.086) \end{bmatrix}$	$\log E_e \longleftarrow \log M3$
49	lag 1 $\begin{bmatrix} -0.790 & 0.200 \\ (0.077) & (0.087) \\ 0 & -0.823 \\ & (0.081) \end{bmatrix}$	$\log E_e \longleftarrow \log M3$
50	lag 1 $\begin{bmatrix} -0.795 & 0.191 \\ (0.076) & (0.086) \\ 0 & -0.821 \\ & (0.081) \end{bmatrix}$	$\log E_e \longleftarrow \log M3$
51	lag 1 $\begin{bmatrix} -0.801 & 0.205 \\ (0.072) & (0.085) \\ 0 & -0.831 \\ & (0.078) \end{bmatrix}$	$\log E_e \longleftarrow \log M3$
52	lag 1 $\begin{bmatrix} -0.810 & 0.218 \\ (0.068) & (0.084) \\ 0 & -0.836 \\ & (0.076) \end{bmatrix}$	$\log E_e \longleftarrow \log M3$

<sup>a</sup> Using the GLS estimation procedure.

<sup>b</sup> Standard errors in parentheses.

<sup>c</sup> In the pattern  $w \longrightarrow z$ : w Granger causes z.



## 5.5 Detecting Granger-causality in the Hong Kong Stock Market

In this section, causality between the Euro and the Hong Kong stock market is examined. The identification approach of ZNZ patterned VAR modelling as proposed in Chapter 3 is utilised to conduct the selection of the optimal ZNZ patterned model. Since there is a huge number of candidate ZNZ patterned models, the search algorithm proposed in Penm and Terrell (1984a) is carried out. This search algorithm employs a block Choleski decomposition in conjunction with model selection criteria to select the optimal patterned VAR without evaluating all possible candidate models. The optimal model is then used as a basis for detecting causal relations among the variables.

To assess the Euro's implications for the Hong Kong stock market, a maximum order of 36 is assigned to the vector system described in Section 5.3, and the search algorithm is undertaken to obtain the optimal ZNZ patterned VAR model. Each of three order selection criteria - Akaike, Schwarz and Hannan - is used to determine the best specification. The ability of these order selection criteria to determine the true specification of a stationary VAR has been examined using a simulation approach suggested by Penm and Terrell (1984b). Their results suggest that SC is superior in order-identification to the other two alternatives in ZNZ patterned VAR modelling for causality studies. Therefore only the specification determined by SC is emphasised and used as the benchmark model for analysing lead-lag relations.

The coefficient estimates of the chosen specification using the adjusted Yule-Walker relations are presented in Table 5.3. To check the adequacy of the model fit, the strategy suggested in Tiao and Tsay (1989) is used, with the proposed algorithm applied to test the residual vector series, using the SC criterion.<sup>19</sup> The results in Table 5.3 support the residual vector being a white noise process. The procedures outlined in Section 3.4 to obtain the GLS-YW estimator are then carried out, with the resultant comparative output also presented in Table 5.3. The detected causal pattern and relationships are presented in Tables 5.4 and 5.5 respectively.

---

<sup>19</sup> Tiao and Tsay (1989) proposed an algorithm using the  $\text{crit}(m,j)$  criterion to select the vector autoregressive moving average process with zero entries. After the final model is selected, their algorithm was then applied to the residual series to test whether this series is a vector white noise process.

**Table 5.3**  
**The optimal ZNZ patterned VAR selected by SC<sup>a,b</sup>**

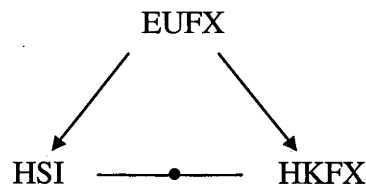
$$y(t) = \{\log \text{EUFX}, \log \text{HSI}, \log \text{HKFX}\}'$$

Maximum order assigned for search	36					
Order of the optimal VAR selected	1					
Coefficient estimator	LS			GLS		
Type of coefficient matrices selected	$\begin{bmatrix} -0.963 & 0 & 0 \\ (0.017) & & \end{bmatrix}$			$\begin{bmatrix} -0.963 & 0 & 0 \\ (0.017) & & \end{bmatrix}$		
$\hat{A}_1$	$\begin{bmatrix} -0.168 & -0.942 & 0 \\ (0.059) & (0.018) & \\ 0.001 & 0 & -0.870 \\ (0.0004) & & (0.031) \end{bmatrix}$			$\begin{bmatrix} -0.006 & -0.941 & 0 \\ (0.002) & (0.017) & \\ 0.005 & 0 & -0.870 \\ (0.002) & & (0.030) \end{bmatrix}$		
Estimate of residual variance-covariance matrix ( $\times 10^{-4}$ )	$\begin{bmatrix} 0.3783 & 0.1239 & -0.0007 \\ 0.1239 & 2.732 & 0.0006 \\ -0.0007 & 0.0006 & 0.00009 \end{bmatrix}$					

<sup>a</sup>SC is also applied to the residual vector. The results support the hypothesis that the residual vector is a white noise process.

<sup>b</sup>The values in parentheses are standard errors of the non-zero coefficient estimates.

**Table 5.4**  
**Causal pattern detected in the three variable system selected by SC<sup>a</sup>**



<sup>a</sup>  $x \rightarrow y$  denotes that  $x$  Granger-causes  $y$  only and not instantaneously;

$x \text{ --}\bullet\text{--} y$  denotes that no causal relation between  $x$  and  $y$ .

**Table 5.5**  
**Causal relationships in the three variable system selected by SC**

Caused by the following variables	EUFX	HSI	HKFX
EUFX	-	one-day causal effect	one-day causal effect
HSI	nil	-	nil
HKFX	nil	nil	-

The relationships identified by the three selection criteria are markedly similar. All the determined specifications consistently indicate that Euro exchange rate is a significant variable that provides leading information for other components of the system. The lagged Euro exchange rate enters not only its own equation but also those of HSI and HKFX, which indicate respectively the Hang Seng stock index and the Hong Kong Dollar relative to the US Dollar in Section 5.3. In all the determined specifications, the lagged level of HSI does not enter any of the exchange rate equations, indicating that variations in the Hong Kong stock market index provide little leading information for the Euro. Also no lagged HKFX components enter the equation of the HSI and EUFX, indicating local forex contains little leading information for either the stock market or the Euro. These results are consistent with economic intuition.

However, it is feasible that the Euro could lead the Hong Kong market. Recall that Hong Kong is an open market and capital flows to and from Europe constitute a component of trading in the Hong Kong stock market. Over the sample period, the Hong Kong stock market rose almost 90 percent while the Euro depreciated. It is reasonable to suspect that these conditions have contributed to a flow of capital between the European and Hong Kong markets.

The situation is undoubtedly far more complex than the above. For instance it may be that the overnight Euro market provides a signal, or reflects, global influences that in turn manifest themselves in the next day's performance of the Hong Kong market. A more complete analysis would include other economic and financial variables such as net capital flows, interest rates and money supply, which could all play a significant role in these markets. Indeed, the model could be extended to incorporate the recent work of Bekaert *et al* (1999) who propose a larger system.

## **5.6 Summary**

This chapter examines two issues in the context of ZNZ patterned VAR modelling. The first issue concerns the causal relationships between movements of the Euro's exchange rate and the money supply, while the second examines the relationship between the Hong Kong stock market and foreign exchange markets.

First, the results show that money supply shocks contribute to movements of the Euro exchange rate, but no causal relationship is detected from the Euro to money supply. These findings are consistent with the standard theory which proposes that an expansion /contraction in monetary policy is associated with a decrease/increase in domestic interest rates for a given expected inflation rate, thus leading to a depreciation/appreciation of the domestic currency.

Second, the findings indicate that movements in the Euro are related to movements in the Hong Kong market, particularly the Hong Kong Dollar. A shock to the Euro foreign exchange market impacts on the movements of both local Hong Kong stock and forex markets. However a shock to either the local stock market or the local forex yields no response from other components of the system. These findings confirm that Hong Kong is susceptible to external shocks.

## CHAPTER 6

### PURCHASING POWER PARITY TESTS IN FOREIGN EXCHANGE MARKETS

#### 6.1 Introduction

The PPP hypothesis implies that, in the long run, changes in the exchange rate between the currencies of two countries reflect changes in the ratio of those countries' price levels. While empirical testing of the PPP hypothesis has received significant attention, the introduction of unit root tests and cointegration theory has renewed interest in this topic. As described in Chapter 2, a wide range of theoretical and empirical models have been built around PPP. However empirical tests have provided inconclusive evidence of its validity.

A major criticism of the classical tests for non-stationarity, such as the augmented Dickey-Fuller test [see Dickey and Fuller (1979)], the Phillips Z test [see Phillips (1987)] and the Phillips-Perron test [see Phillips and Perron (1988)], is that they lack power to distinguish between unit root processes and near-unit root stationary processes [see Enders (1995) and Harris (1995)]. Therefore, they have a tendency to accept the null hypothesis of non-stationarity [see Hakkio (1986) and DeJong *et al* (1989)]. This feature has prompted the use of tests which employ the null hypothesis of stationarity [see Fisher and Park (1991) and Kwiatkowski *et al* (1992)].

International experience has demonstrated that the tests for PPP are sensitive to the null hypothesis employed. For instance, PPP can be rejected under a null hypothesis embodying the presence of a unit root, but accepted in a test with a null hypothesis of stationarity. This chapter demonstrates the sensitivity of PPP testing to the nature of the unit root tests. Three unit root tests are applied to fourteen real bilateral exchange rates. The first two, the augmented Dickey-Fuller test and the Phillips-Perron test, are procedures using the null hypothesis of a unit root. The third, the Kwiatkowski *et al* test, is a procedure which employs the null hypothesis of stationarity. In future analysis the necessary condition and the necessary and sufficient condition for PPP are sequentially tested for fourteen bilateral exchange rates. This test is undertaken in the framework of subset vector error-correction modelling (VECM) with zero coefficients. Since VECM modelling can accommodate both long-term and dynamic responses, this approach is different from scalar unit root based methods. Of the fourteen exchange rates tested, support for the necessary condition for PPP is found in half of them. The necessary and sufficient condition for PPP is then tested using both a bootstrap procedure and an F test. This condition is consistently accepted for three of the seven exchange rates.

The remainder of this chapter is organised as follows. Section 6.2 describes the data. Section 6.3 examines bilateral exchange rates using unit root tests. Section 6.4 investigates PPP by using subset VECMs with zero coefficients. Section 6.5 tests for the necessary and sufficient condition for PPP. Section 6.6 examines the PPP conditions in the Australian foreign exchange market using an I(2) analysis



in a three-variable ZNZ patterned VECM framework. Section 6.7 presents a brief summary to conclude the chapter.

## **6.2 Data Description**

Data are obtained from the International Monetary Fund through the Statistical Analysis and Retrieval Service (STARS), maintained by the Australian-Japan Research Centre at the Australian National University. The sample covers quarterly data from 1975(1) to 1994(4). The exchange rate series is the quarterly average of the domestic currency per unit of the US Dollar. In the cases of Japan, Germany, Australia, Spain, Korea, Indonesia, Singapore, the Philippines and Thailand, the wholesale price index (WPI) was used to approximate domestic and foreign price levels. In the cases of France, the United Kingdom, Italy, Hong Kong and Malaysia, the consumer price index (CPI) was used (as a result of data availability). The US WPI was used to approximate the foreign price level.

Theoretically, the WPI is preferred in the test for PPP as the proportion of non-traded goods is lower in the WPI than that in the CPI. In essence, if WPI is used, the relationship being tested is closer to the 'law of one price' for traded goods. However, with many non-traded goods and services included in the CPI, the test results using CPI would be dependent on whether the relative price of traded to non-traded goods has changed more in one country than in the other over the sample period. Kim (1990) has examined the use of CPI and WPI in the test for PPP for a number of industrial countries including Canada, France, Italy, Japan

and the United States. His results indicate that, in some cases, the hypothesis of PPP is found to be acceptable using the WPI, but not using the CPI.

Table 6.1 shows the exchange rate regimes for these countries. There have been some changes to the regimes of the Asian exchange rate (excluding Japan) over the sample period. These changes are also presented in Table 6.1. The Australian Dollar switched from a fixed exchange rate regime to a flexible one in December 1983.

**Table 6.1****Fixed and flexible exchange rate regimes, and the Exchange Rate Mechanism**

<b>Country</b>	<b>Period</b>	<b>Exchange rate regime</b>
South Korea	1980 to 1994	Flexible
	prior to 1980	Fixed
Thailand	prior to 1984	Fixed
	1984 to 1995	Fixed
Singapore	1987 to 1994	Flexible
	prior to 1987	Fixed
Hong Kong	1983 to 1994	Fixed
Malaysia	1978 to 1994	Fixed
Indonesia	1978 to 1994	Flexible
	prior to 1978	Fixed
Philippines	1975 to 1994	Flexible
United States	1975 to 1994	Flexible
France	1979 to 1994	Exchange Rate Mechanism
Germany	1979 to 1994	Exchange Rate Mechanism
Japan	1975 to 1994	Flexible
Spain	1979 to 1994	Exchange Rate Mechanism
United Kingdom	1992 to 1994	Flexible
	1990 to 1992	Exchange Rate Mechanism
	prior to 1990	Flexible
Italy	1979 to 1994	Exchange Rate Mechanism
Australia	1983 to 1994	Flexible
	prior to 1983	Fixed

Source: International Monetary Fund

### 6.3 Purchasing Power Parity and Unit Root Tests

As described in Section 2.4, there is an extensive literature that traverses the methods of unit root tests and cointegration tests for PPP. A general outcome of these studies is that long-term PPP appears not to hold, when tests based on short- or medium-length time-series are used [see Roll (1979), Mishkin (1984), and Piggot and Sweeney (1985)], but appears to hold in longer time samples [see Abuaf and Jorion (1990), Froot and Rogoff (1994), and Lothian and Taylor (1996)]. This is because statistical tests become less powerful in small samples. Another consensus result is that higher-frequency data (for example monthly data) may not yield evidence of PPP in the long-term [see McNown and Wallace (1989), Taylor (1985), and Corbae and Ouliaris (1988)]. However when researchers [see Edison (1987) and Kim (1990)] shift to low-frequency data and use cointegration techniques to test the PPP, the evidence usually supports the long-term convergence of real exchange rates toward PPP.

More recently, VECM models have given an opportunity to develop a more complex process in financial markets. These models can accommodate both long-term and dynamic responses. Balancing these arguments can be achieved through the use of a larger time-series. Given the availability of data, we use quarterly data in the following tests.

To demonstrate the sensitivity of PPP to the test procedure used three unit root tests are employed. The first two, the augmented Dickey-Fuller test [Dickey and Fuller (1979)] and the Phillips-Perron test [Phillips and Perron (1988)], are

procedures using the null hypothesis of a unit root. The third, the Kwiatkowski *et al* (1992) test, is a procedure which employs the null hypothesis of stationarity. Each test was applied to the generated real bilateral exchange rates between the United States and fourteen other major OECD and Asian economies. The test results are presented in Table 6.2.

Of the fourteen real exchange rates, the PPP hypothesis is consistently rejected by all three unit root tests for the Japanese Yen, the Spanish Peseta and the Thai Baht. The PPP hypothesis is consistently accepted by all three unit root tests for the Philippine Peso. For the remaining ten exchange rates, the PPP hypothesis is consistently rejected by both the Dickey-Fuller test and the Phillips-Perron test with the null hypothesis of a unit root, but consistently accepted by the Kwiatkowski *et al* test under the null hypothesis of stationarity.

PPP is consistently accepted or rejected for only four exchange rates by all three tests, and is inconsistent across the other ten exchange rates. This inconsistency is a major problem in interpreting and making conclusions. As these tests only examine stationarity or non-stationarity in the residuals, these tests impose restrictions before testing the necessary condition for PPP. In the remainder of the chapter, a new test procedure is introduced for PPP. The procedure is tested in the framework of subset VECM with zero coefficients. Since this approach can accommodate both long-term and dynamic responses, it is different from those based on unit root tests.

**Table 6.2**  
**Test results for purchasing power parity of bilateral exchange rates from**  
**1975(1) to 1994(4) using unit root tests**

Country	Dickey-Fuller test	Phillips- Perron test	Kwiatkowski <i>et al</i> test
<i>OECD exchange rates</i>			
Japan	-1.28*	-1.98*	0.741*
Germany	-1.82	-1.80	0.313
France	-2.13	-1.68	0.188
The United Kingdom	-2.22	-2.15	0.165
Italy	-2.16	-2.08	0.362
Spain	-1.65*	-1.88*	0.812*
Australia	-1.61	-1.96	0.119
<i>Asian exchange rates</i>			
Singapore	-1.33	-1.85	0.134
Hong Kong	-0.62	-0.98	0.330
Korea	-2.16	-2.14	0.273
Indonesia	-2.46	-2.55	0.134
Thailand	-1.92*	-1.44*	0.508*
Malaysia	-0.93	-1.47	0.128
Philippines	-4.03**	-3.58**	0.099**

The results of both the augmented Dickey-Fuller test and the Phillips-Perron test are based on four lag terms, while the results of the Kwiatkowski *et al* (1992) test are based on the inclusion of six lag truncation parameters. Lag choice appears to have little impact on the reported results. The symbol \* denotes that all three unit root tests consistently reject the PPP hypothesis tested at the 5 percent level of significance, and \*\* denotes that all three unit root tests consistently accept the PPP hypothesis tested at the 5 percent level.

## 6.4 Test for PPP Using Subset VECMs with Zero Coefficients

In this section the Penm *et al* (1997) approach is applied to the individual VECMs formed by nominal exchange rates and the price ratios. The approach provides an effective search algorithm, in conjunction with a model selection criterion, to determine the optimal subset specification with zero coefficients for a VECM. To implement this procedure, AIC<sup>20</sup> is used for model selection.

Consider a generalised VECM expressed as follows.

$$A^p(1)y_{t-1} + A^{p-1}(L)\Delta y_t = \varepsilon_t, \quad (6.1)$$

where  $y_t$  contain  $I(1)$  variables,  $\Delta$  denotes first difference and

$$A^{p-1}(L) = I + \sum_{\tau=1}^{q-1} A_{\tau}^* L^{\tau}, \quad L \text{ is the lag operator and the roots of } |A^p(L)| = 0 \text{ lie outside}$$

or on the unit circle.

If  $y_t$  is cointegrated, then the long-term impact matrix,  $A^p(1)$ , must be singular.

As a result,  $A^p(1) = \alpha\beta'$ , where the columns of  $\beta$  are the cointegrating vectors and the rows of  $\alpha$  are the loading vectors.  $A^{p-1}(L)\Delta y_t$  is called the VAR part of the VECM.

---

<sup>20</sup> The general approach of this thesis is to carefully examine different characteristics of criteria, and normally HQC or SC is chosen. However in comparing work conducted in this area [Cheung and Lai (1993), Cheng (1999)], the AIC criterion has been commonly used in testing the PPP hypothesis. Therefore for comparison purposes, the AIC is used in this case.

The strength of the Penm *et al* procedure is that it allows a subset structure with zero coefficients to be incorporated in error-correction modelling. If the long-term impact matrix,  $A^p(1)$ , contains zero coefficients, then the cointegrating and loading vectors may also contain zero coefficients. This inclusion enhances the modelling power of ZNZ patterned VECMs, especially for finite samples.

The Penm *et al* procedure is applied to the VECMs formed by individual bilateral nominal exchange rates and the relevant price ratios. The necessary condition for each case is examined through the determined ZNZ pattern for the long-term impact matrix,  $A^p(1)$ . If the determined  $A^p(1)$  is a singular matrix, then the nominal exchange rate and the price ratio are cointegrated. Consequently the necessary condition for PPP is accepted.

Compared to OECD countries, most Asian economies have been subject to major structural changes due to trade restrictions, resource controls and government intervention, or even government instability. Therefore these changes cause increased political risks<sup>21</sup> [see Mahoney *et al* (2001)]. Given this level of change,

---

<sup>21</sup> Political risks arise from changes in the political environment that may adversely affect the value of a firm's business activities. Political risk is an important component in the capital budgeting process for foreign direct investment. It can affect asset prices, and thus the movements of exchange rates. Bekaert and Harvey (1998) examine the impact of capital market liberalisations on various emerging markets. They find a reduction in the cost of capital after market liberalisation. Cherian and Perotti (2001) investigate asset prices in a context of uncertainty about future government policy. They reveal that, as current policy is maintained, perceived risk falls. This will lead to a gradual appreciation of asset prices and a gradual decrease in their conditional variance.



this thesis has an *a priori* reason to believe that there may be larger deviations from their long-term PPP level.

Previous work [see Cheung and Lai (1993)] suggests that the lack of support for the necessary condition for PPP could be driven by the existence of non-traded goods and services, and by measurement problems of consumer price series. As pointed out by Cheung and Lai, the price-level measurement problems associated with index construction and aggregation can result in the rejection of PPP. Other measurement problems such as international differences and variations in product qualities and consumption patterns, and between transaction and listed prices, can also weaken the relationship between the ratios of price levels and exchange rates.

Of the seven OECD exchange rates tested, the necessary condition for PPP is accepted for the German Mark, the French Franc, the Italian Lira and the Australian Dollar. This condition, however, is rejected for the Japanese Yen, the UK Pound and the Spanish Peseta. The determined VECMs for those OECD exchange rates which satisfy the necessary condition are presented in Table 6.3.

**Table 6.3**

**Estimated VECMs which support the necessary condition for PPP of OECD exchange rates**

Country	Model
Germany	$\begin{bmatrix} d \ln E_t \\ d \ln(P_G/P_{US})_t \end{bmatrix} = \begin{bmatrix} 0.293 & 0 \\ (2.75) & \\ 0 & 0.305 \\ & (2.96) \end{bmatrix} \begin{bmatrix} d \ln E_{t-1} \\ d \ln(P_G/P_{US})_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.022 & -0.060 \\ (3.37) & (3.92) \end{bmatrix} \begin{bmatrix} \ln E_{t-1} \\ \ln(P_G/P_{US})_{t-1} \end{bmatrix}$
France	$\begin{bmatrix} d \ln E_t \\ d \ln(P_F/P_{US})_t \end{bmatrix} = \begin{bmatrix} 0.325 & 0 \\ (3.07) & \\ 0 & 0.184 \\ & (1.94) \end{bmatrix} \begin{bmatrix} d \ln E_{t-1} \\ d \ln(P_F/P_{US})_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.073 & -0.026 \\ (5.49) & (5.04) \end{bmatrix} \begin{bmatrix} \ln E_{t-1} \\ \ln(P_F/P_{US})_{t-1} \end{bmatrix}$
Italy	$\begin{bmatrix} d \ln E_t \\ d \ln(P_I/P_{US})_t \end{bmatrix} = \begin{bmatrix} 0.364 & 0 \\ (3.45) & \\ 0 & 0.290 \\ & (2.95) \end{bmatrix} \begin{bmatrix} d \ln E_{t-1} \\ d \ln(P_I/P_{US})_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0.785 \\ & (1.63) \\ 0.047 & 0 \\ (2.45) & \end{bmatrix} \begin{bmatrix} d \ln E_{t-2} \\ d \ln(P_I/P_{US})_{t-2} \end{bmatrix}$ $+ \begin{bmatrix} 0.224 & 0 \\ (2.06) & \\ 0 & 0.188 \\ & (2.01) \end{bmatrix} \begin{bmatrix} d \ln E_{t-3} \\ d \ln(P_I/P_{US})_{t-3} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0.242 \\ & (2.24) \end{bmatrix} \begin{bmatrix} d \ln E_{t-4} \\ d \ln(P_I/P_{US})_{t-4} \end{bmatrix}$ $- \begin{bmatrix} 0.103 & -0.101 \\ (2.55) & (2.18) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \ln E_{t-1} \\ \ln(P_I/P_{US})_{t-1} \end{bmatrix}$
Australia	$\begin{bmatrix} d \ln E_t \\ d \ln(P_A/P_{US})_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0.029 & -0.079 \\ (2.86) & (3.69) \end{bmatrix} \begin{bmatrix} \ln E_{t-1} \\ \ln(P_A/P_{US})_{t-1} \end{bmatrix}$

t-statistics in brackets. d denotes first difference.  $E_t$  denotes the units of domestic currency per unit of the US Dollar,  $P_{US}$  the US price level and  $P_i$  the price level for country i, i = Germany, France, Italy and Australia. The estimation was undertaken using the GLS method.

Of the seven Asian exchange rates, this condition is found accepted for the Indonesian Rupiah, the Singapore Dollar and the Philippine Peso. The determined subset VECMs, which allow for possible zero coefficients are presented in Table 6.4. The procedure developed by Penm and Terrell (1984a) is applied to the vector autoregressive process formed by the residuals of each estimation. The results indicate that autocorrelation in the residuals is not a problem.

**Table 6.4**

**Estimated VECMs which support the necessary condition for PPP of Asian  
exchange rates**

Country	Model
Indonesia	$\begin{bmatrix} d \ln E_t \\ d \ln(P_I/P_{US})_t \end{bmatrix} = \begin{bmatrix} 0.472 & 0 \\ (4.89) & 0.761 \\ 0 & (11.45) \end{bmatrix} \begin{bmatrix} d \ln E_{t-1} \\ d \ln(P_I/P_{US})_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.009 & -0.023 \\ (2.54) & (2.49) \end{bmatrix} \begin{bmatrix} \ln E_{t-1} \\ \ln(P_I/P_{US})_{t-1} \end{bmatrix}$
Singapore	$\begin{bmatrix} d \ln E_t \\ d \ln(P_S/P_{US})_t \end{bmatrix} = \begin{bmatrix} 0 & -0.167 \\ (1.69) & 0.384 \\ 0 & (3.44) \end{bmatrix} \begin{bmatrix} d \ln E_{t-1} \\ d \ln(P_S/P_{US})_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -0.304 \\ (2.75) & \end{bmatrix} \begin{bmatrix} d \ln E_{t-2} \\ d \ln(P_S/P_{US})_{t-2} \end{bmatrix} +$ $\begin{bmatrix} 0 & -0.220 \\ (2.40) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} d \ln E_{t-4} \\ d \ln(P_S/P_{US})_{t-4} \end{bmatrix} - \begin{bmatrix} 0.047 & -0.051 \\ (2.35) & (2.11) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \ln E_{t-1} \\ \ln(P_S/P_{US})_{t-1} \end{bmatrix}$
Philippines	$\begin{bmatrix} d \ln E_t \\ d \ln(P_P/P_{US})_t \end{bmatrix} = \begin{bmatrix} 0.420 & 0 \\ (4.00) & 0.237 & 0.247 \\ (2.32) & (2.56) \end{bmatrix} \begin{bmatrix} d \ln E_{t-1} \\ d \ln(P_P/P_{US})_{t-1} \end{bmatrix} + \begin{bmatrix} 0.232 & 0 \\ (2.18) & 0.218 & 0 \\ (2.63) & \end{bmatrix} \begin{bmatrix} d \ln E_{t-3} \\ d \ln(P_P/P_{US})_{t-3} \end{bmatrix} +$ $\begin{bmatrix} 0 & 0 \\ 0.171 & -0.158 \\ (2.72) & (2.75) \end{bmatrix} \begin{bmatrix} \ln E_{t-1} \\ \ln(P_P/P_{US})_{t-1} \end{bmatrix}$

t-statistics in brackets. d denotes first difference.  $E_t$  denotes the units of domestic currency per unit of the US Dollar,  $P_{US}$  the US price level and  $P_i$  the price level for country i, i = Indonesia, Singapore and The Philippines. The estimation was undertaken using the GLS method.

The VECMs presented in Tables 6.3 and 6.4 exhibit some interesting features on the interrelationship between the nominal exchange rate and the price ratio. First, in each case, the ZNZ pattern determined for the long-term impact matrix,  $A^q(1)$ , can be used to obtain the ZNZ patterns for  $\alpha$  and  $\beta$ . For example, in the case of the Italian Lira, the ZNZ pattern of  $A^q(1)$  is  $\begin{bmatrix} 0.103 & -0.101 \\ 0 & 0 \end{bmatrix}$ . From this pattern, both  $\alpha = \begin{bmatrix} 0.103 \\ 0 \end{bmatrix}$  and  $\beta' = [1 \quad -0.98]$  are obtained. The determined ZNZ patterns of  $\alpha$  and  $\beta$  indicate that, in this case, the nominal exchange rate and the price ratio are cointegrated, and the VAR part of the VECM is useful in explaining the short-term variations in the nominal exchange rate. Similar conclusions can be drawn for the Singapore Dollar. In other cases presented in Tables 6.3 and 6.4, the determined VECMs are found useful in explaining the short-term variations in the ratio of domestic to foreign prices. Specifically, these cases are the German Mark, the French Franc, the Australian Dollar, the Indonesian Rupiah and the Philippine Peso.

Second, a simple lag structure is determined for the VAR part of the VECM for the German Mark, the French Franc, the Australian Dollar and the Indonesian Rupiah. This indicates that, for these exchange rates, the determined VECMs provide little information about the short-term variations between the nominal exchange rate and relative prices. However, it is a different case for the remaining exchange rates, namely the Italian Lira, the Singapore Dollar and the Philippine Peso. The VECMs determined for these exchange rates present a more complex structure for the short-term interaction between the nominal exchange rate and the ratio of domestic to foreign prices.

The above outcomes provide sound evidence in favour of long-term PPP, in contrast to previous studies [see Hakkio (1984)].

### 6.5 Test for the Necessary and Sufficient Condition

To test for the necessary and sufficient condition, that is  $\beta' = (1, -1)$ , two approaches are utilised. One is an F test and the other is a bootstrap procedure [see Hall (1992)]. These procedures are both undertaken in the framework of VECMs. Because the VECMs presented in Tables 6.3 and 6.4 are stationary, standard asymptotic results for hypothesis testing apply. Under the null hypothesis,  $H_0: \beta' = (1, -1)$ , the corresponding coefficients of the long-term impact matrix,  $A^q(1)$ , must have the same value but opposite signs. For example, in the case of the Italian Lira, if  $\beta' = (1, -1)$ , then  $A^q(1) = \begin{bmatrix} y & -y \\ 0 & 0 \end{bmatrix}$ . Therefore, the necessary and sufficient condition for PPP can be examined through hypothesis testing of the associated coefficient estimates of the long-term impact matrix.

The F test is undertaken under the null hypothesis that the coefficient estimates of the long-term impact matrix have the same value but opposite signs. The test results are presented in Table 6.5. Of the seven exchange rates for which the necessary condition is found acceptable, the necessary and sufficient condition cannot be rejected for the Italian Lira, the Singapore Dollar and the Philippine Peso.

**Table 6.5**  
**Test of the necessary and sufficient condition for PPP**

Country	<u>Bootstrap Probability</u>	<u>F-test</u>
Germany	11%	9.84*
France	12%	24.77*
Italy	82%	0.02
Australia	15%	14.41*
Indonesia	29%	5.45*
Singapore	85%	0.06
Philippines	90%	3.45

The bootstrap results are based on 1000 replications. The critical value for the F test at the 5 percent level is 3.84. The symbol \* denotes that the necessary and sufficient condition for PPP is rejected at the 5 percent level.

In an attempt to establish some consistency for these test results, a bootstrap procedure is also utilised to examine the necessary and sufficient condition for PPP. In this procedure, confidence intervals for the coefficient estimates of the VECM are first developed using the bootstrap method. The procedure can then calculate the probability of the associated coefficient estimates of the long-term impact matrix having the same value but opposite signs as implied by the PPP hypothesis. The bootstrap results are consistent with those of the F test. Of the seven exchange rates examined, the Italian Lira, the Singapore Dollar and the Philippine Peso again exhibit the highest probability of accepting the necessary and sufficient condition for PPP.

One further finding is potentially noteworthy. The VECMs for the three exchange rates that accept the necessary and sufficient condition are found to exhibit a more complex lag structure than those for which this condition is rejected. Since the test for the necessary and sufficient condition depends upon hypothesis testing of the coefficient estimates of the long-term impact matrix, these results suggests that the test for the necessary and sufficient condition could be influenced by the short-term lag structure in the system.

## 6.6. An I(2) Analysis of PPP

In Section 6.4 the ratio  $\log(\frac{P_t}{P_t^*})$  is treated as one variable, which does not necessarily have the same integration order as  $\log(P_t)$  and  $\log(P_t^*)$  [see Corbae and Ouliaris (1990) and Oh (1996)]. If the three variables are specified as  $\log E_t$ ,  $\log P_t$  and  $\log P_t^*$ , the order of integration of each variable may not be same, that is  $\log E_t$  is I(1), but both  $\log P_t$  and  $\log P_t^*$  are often found to be I(2). Therefore this could suggest a need for an I(2) model, rather than an I(1) model.

In cointegration theory Granger and Lee (1989) have suggested multi-cointegration to improve short- and long-term forecasts. Engle and Yoo (1991) have proposed an I(2) cointegration system which coincides with Granger's multi-cointegration. Diamandis *et al* (2000) have conducted an I(2) analysis to examine the long-term properties of the monetary exchange rate model, under the hypothesis that the system contains I(2) variables.



An  $I(2)$  algorithm for selecting cointegrating and loading vectors in a ZNZ patterned VECM for PPP testing is developed in this section. Section 6.6.1 proposes procedures for identifying the optimal specification for a ZNZ patterned VECM for an  $I(2)$  system. After the optimal ZNZ VECM is identified, the rank of the long-term impact matrix is then computed using the singular value decomposition method, such that the number of cointegrating vectors in the system is known. Section 6.6.2 introduces a tree-pruning algorithm for the search of all acceptable ZNZ patterns of the cointegrating and loading vectors. The estimation of the associated candidates for the ZNZ patterned loading vectors in the VECM framework is then carried out by the regression method with linear restrictions as proposed in Penm *et al* (1997). Section 6.6.3 examines the PPP conditions for the Australian foreign exchange markets in a three-variable ZNZ VECM for an  $I(2)$  system. Section 6.6.4 deals with a three-variable system concerning the stock market.

### 6.6.1 VECM Modelling for an $I(2)$ System

As described in Section 4.3,  $y(t)$  is integrated of order  $d$ ,  $I(d)$ , if it contains at least one element which must be differenced  $d$  times before it becomes  $I(0)$ . This thesis also calls  $y(t)$  cointegrated with the cointegrating vector,  $\beta$ , of order  $g$ , if  $\beta'y(t)$  is integrated of order  $(d-g)$ , where  $y(t)$  has to contain at least two  $I(d)$  variables.<sup>22</sup>

---

<sup>22</sup> In this section only the case  $d=2$  is considered, although the procedure can be generally applied to models with  $d>2$ .

The following decompositions are always valid mathematically:

$$\begin{aligned} A^p(L) &= A^p(1)L + (I - L)A^{p-1}(L) \\ &= A^p(1)L + A^{p-1}(1)L - A^{p-1}(1)L^2 + (I - L)^2 A^{p-2}(L) \end{aligned}$$

Under the assumption of an I(2) system we use the decomposition:

$$A^p(L) = A^p(1)L + A^{p-1}(1)L(I - L) + A^{p-2}(L)(I - L)^2.$$

Following Engle and Yoo (1991), the equivalent VECM for an I(2) system can be expressed as:

$$\begin{bmatrix} A^p(1) & , & A^{p-1}(1) \end{bmatrix} \begin{bmatrix} y(t-1) \\ \Delta y(t-1) \end{bmatrix} + A^{p-2}(L)\Delta^2 y(t) = \varepsilon(t), \quad (6.2)$$

where  $y(t)$  contains variables of three types, namely I(0), I(1) and I(2) and

$\Delta = (I - L)$ . (6.2) can be rewritten as:

$$A^* \begin{bmatrix} y(t-1) \\ \Delta y(t-1) \end{bmatrix} + A^{p-2}(L)\Delta^2 y(t) = \varepsilon(t), \quad (6.3)$$

where  $A^* = \begin{bmatrix} A^p(1) & , & A^{p-1}(1) \end{bmatrix}$ ,  $A^* \begin{bmatrix} y(t-1) \\ \Delta y(t-1) \end{bmatrix}$  is stationary and the error correction

term. The term  $A^{p-2}(L)\Delta^2 y(t)$  is the vector autoregressive part of the VECM.

Further, it is necessary to consider a hypothesis where every  $(i,j)$ -th element, for specified  $i$  and  $j$ , is zero in all coefficient matrices in a VAR. If this hypothesis is framed in the VAR, it can be described as:

$$\sum_{k=0}^p A_k y(t-k) = A^p(L)y(t) = \varepsilon(t), \quad (6.3.a)$$

which is equation (3.1) in Chapter 3. These zero entries will also hold in the error-correction terms and in the vector autoregressive part of the equivalent VECM for an  $I(2)$  system, say (6.2).

Analogously this thesis can achieve a result that if all  $(i,j)$ -th coefficient elements in the error-correction terms and all  $(i,j)$ -th coefficient elements in the vector autoregressive part of this VECM are zeros, then every  $(i,j)$ -th entry is zero for all coefficient matrices in a VAR.

The implications of the above outcome are straightforward. If  $y_j$  does not Granger-cause  $y_i$ , then any  $(i,j)$ -th entry must be zero for all coefficient matrices in the VAR. Also all  $(i,j)$ -th coefficient elements in the equivalent VECM for an  $I(2)$  system are zeros.

In a similar way, it can be demonstrated that if  $y_j$  does Granger-cause  $y_i$ , then the  $(i,j)$ -th element of  $A^p(L)$  in (6.3.a) is nonzero. Also, at least a single  $(i,j)$ -th coefficient element is nonzero in  $A^p(1)$ ,  $A^{p-1}(1)$  or  $A^{p-2}(L)$  in the equivalent VECM. Of note, an indirect causality from  $y_j$  to  $y_i$  through  $y_m$  indicates  $y_j$  causing  $y_i$  but only through  $y_m$ . Hence,  $y_j$  Granger-causes  $y_m$ ,  $y_m$  Granger-

causes  $y_i$ , and  $y_j$  does not Granger-cause  $y_i$  directly. It can be easily demonstrated that the VAR in (3.1) has nonzero  $(m,j)$ -th and  $(i,m)$ -th elements and a zero  $(i,j)$ -th element in  $A^p(L)$ . This indirect causality can also be shown in the equivalent VECM, which has at least a single nonzero  $(m,j)$ -th element and a single nonzero  $(i,m)$ -th elements in  $A^p(1)$ ,  $A^{p-1}(1)$  and  $A^{p-2}(L)$ . Also all the  $(i,j)$ -th elements in the equivalent VECM are zeros.

Thus Granger causality, Granger non-causality and indirect causality can be detected from both the ZNZ patterned VECM and its equivalent ZNZ patterned VAR are identical. Hence the ZNZ patterned VECM is a more straightforward and effective means of testing for the Granger causal relations. The same benefits will be present if the ZNZ patterned VECM is used to analyse cointegrating relations.

### 6.6.2 Search Algorithm for an I(2) System

In the proposed algorithm for an I(2) system, the identification of ZNZ patterned  $A^*$  and the determination of ZNZ patterned  $\alpha$  and  $\beta$  are carried out in the following way. First, model selection criteria are used to select the optimal subset VECM with zero entries to determine the ZNZ patterned  $A^*$ . Penm and Terrell (1984a) have proposed a search method in conjunction with model selection criteria to select the optimal subset VAR with zero entries. This method is now extended to select the optimal subset VECM with zero entries for an I(2) system.

Second, after the ZNZ patterned  $A^*$  is determined, the rank of the matrix  $A^*$  is then computed using the singular value decomposition (SVD) method, and the number of cointegrating vectors in the system will be known.

Third, given the ZNZ patterned  $A^*$  has been determined and the rank of  $A^*$  has been computed, the algorithm then proceeds with the tree-pruning algorithm straightforwardly adapted for an I(2) system to obtain all acceptable ZNZ patterned  $\alpha$ s and  $\beta$ s which are consistent with the ZNZ patterned  $A^*$ . Let  $\alpha_p$  and  $\beta_p$  denote a ZNZ pattern of  $\alpha$  and  $\beta$  respectively and  $A_p$  the ZNZ pattern of  $A^*$ . If the  $(i,j)$ -th entry of the product,  $\alpha_p \beta_p'$  is zero, and the corresponding  $(i,j)$ -th entry of  $A_p$  is also zero, then both  $\alpha_p$  and  $\beta_p$  are acceptable. This tree-pruning algorithm, which avoids the need to evaluate all possible ZNZ patterned  $\alpha$ s and  $\beta$ s, is discussed in Appendix B.

The ZNZ patterns of acceptable  $\alpha$  and  $\beta$  depend on the pattern of  $A^*$  determined earlier by model selection criteria. Of note, the imposition of zero entries on  $\beta$  does not preclude a similar restriction on  $\alpha$ . One example is that if the determined  $A^*$  contains a zero row, such as:

$$A^* = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix},$$

where 1 denotes a non-zero entry.

In this case zero restrictions will have to be imposed on the first row of  $\alpha$ . This is because the pattern of  $A^*$  implies that the cointegrating relations in the system have no influence on the first variable in the system. Noting that the number of zeros in  $\alpha$  and  $\beta$  are not fixed even with a given ZNZ patterned  $A^*$ , many differently patterned  $\alpha$  and  $\beta$  can be obtained using the tree-pruning algorithm.

To obtain the correct specification for  $\alpha$  and  $\beta$ , the algorithm next checks to see whether  $\alpha$  and  $\beta$  can be uniquely obtained by factorising  $A^*$ . If this is possible, the factorisation can be carried out. If it is not possible, the efficient estimation of I(2) cointegrated systems based on a triangular ECM representation [see Stock and Watson (1993)] is employed to estimate  $\beta$ . Since any non-zero entry in  $\beta$  could be normalised as unity, the estimation procedure is repeated with all possible normalisations. Again different normalisations in practice may result in different forecasting performances for the model. The normalisation, which produces the smallest value for model selection, is then selected as the candidate  $\beta$ . After the optimal normalisation is determined for every candidate  $\beta$ , the associated acceptable ZNZ patterned  $\alpha$ s are then estimated in the VECM framework and model selection criteria are employed again to determine the optimal  $\alpha$  and  $\beta$ .

There are two reasons for employing model selection criteria again to determine the optimal  $\alpha$  and  $\beta$ . In the example given above, model selection criteria will help to select between (6.4), (6.5) and (6.6), since the approach of zero-maximisation cannot be used to determine  $\beta$ . In addition, Engle and Granger (1987) have demonstrated that efficiency gains could be obtained in such estimation.

### 6.6.3 PPP Testing between the Australian and US Dollar Using an I(2) Analysis

In this section, an application to PPP testing using an I(2) analysis is presented. Another application, which examines the relationships among the stock market, money supply and inflation, is illustrated in Section 6.6.4.

The quarterly seasonally-adjusted consumer price indices for Australia ( $CPI_{AUS}$ ) and the United States ( $CPI_{US}$ ), and the exchange rates (EXCH) per US Dollar from March 1972 through December 1998 are used. The data are obtained from DataStream<sup>TM</sup>.<sup>23</sup> The  $y$  vector comprises  $\log(CPI_{AUS})$ ,  $\log(CPI_{US})$ , as well as  $\log(EXCH)$ , measured as the value of the Australian Dollar relative to the US Dollar. The Dickey and Pantula (1987) tests indicate that both  $\log(CPI_{AUS})$  and  $\log(CPI_{US})$  are I(2) while  $\log(EXCH)$  is I(1). Hence the issue of an I(2) series arises. The results identified by SC are presented in Table 6.6. In addition, to check the adequacy of the model fit, the strategy suggested in Tiao and Tsay (1989) and Penm *et al* (1997) is used, with the proposed Penm and Terrell (1984a) algorithm applied to test the residual vector series, using the SC criterion. The results in Table 6.6 support the residual vector being a white noise process.

The selected pattern of the cointegrating vector demonstrates some interesting findings. The first selected cointegrating vector indicates that  $\Delta \log(EXCH)$  is stationary. The second selected cointegrating vector confirms that both

---

<sup>23</sup> The calculation of the CPI in the USA changed after 1 January 1999, hence the fourth quarter of 1998 is chosen as the end period.

$\log(\text{CPI}_{\text{AUS}})$  and  $\log(\text{CPI}_{\text{US}})$  are cointegrated with  $\log(\text{EXCH})$ . The same sign occurring in  $\log(\text{EXCH})$  and  $\log(\text{CPI}_{\text{AUS}})$ , as shown in Table 6.6, indicates that, *ceteris paribus*, an increase in  $\text{CPI}_{\text{AUS}}$  leads to a depreciation in the Australian Dollar, and the opposite sign occurring in  $\log(\text{EXCH})$  and  $\log(\text{CPI}_{\text{US}})$  indicates that, *ceteris paribus*, an increase in  $\text{CPI}_{\text{US}}$  leads to a depreciation in the US Dollar. The presence of the long-term cointegrating relationships is consistent with PPP holding within the I(2) system and across the Australian and US exchange market.

In looking for causal relations among the nominal exchange rate, and the domestic and foreign price levels, the VECM selected indicates feedback relations exist between the pair of  $\text{CPI}_{\text{US}}$  and  $\text{CPI}_{\text{AUS}}$ <sup>24</sup> and the pair of  $\text{CPI}_{\text{AUS}}$  and  $\text{EXCH}$ , and one-way causation from  $\text{CPI}_{\text{US}}$  to  $\text{EXCH}$ . Although there is no direct Granger causation from  $\text{EXCH}$  to  $\text{CPI}_{\text{US}}$ , there is however indirect causation from  $\text{EXCH}$  to  $\text{CPI}_{\text{US}}$  via  $\text{CPI}_{\text{AUS}}$ . Hence, shocks to any one of the variables will be transmitted through the system. This in turn offers some insights into the dynamics which have been observed among these variables, despite PPP holding in the long-run [see Abuaf and Jorion (1990)].

---

<sup>24</sup> The causation detected from  $\text{CPI}_{\text{AUS}}$  to  $\text{CPI}_{\text{US}}$  is an unexpected result. However this result may be driven by other underlying and fundamental relationships.

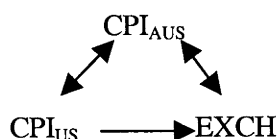


**The VECM<sup>a</sup> for the relationship linking exchange rates and consumer price indices between Australia and the USA**

Long-term Cointegrating Relationship Identified:

- 1) A stationary  $\Delta \log(\text{EXCH})$
- 2)  $\log(\text{EXCH}) = 0.278 \log(\text{CPI}_{\text{AUS}}) - 0.607 \log(\text{CPI}_{\text{US}})$

Granger Causal Pattern<sup>c</sup> Recognised:



- 
- a) Model selected by SC Using the GLS Procedure. Standard errors in parentheses.  $\Delta$  denotes first difference.
- b) For simplicity, the values of SC for  $q > 3$  are not presented, but can be supplied to readers upon request.
- c)  $x$  Granger-causes  $y$ : (Notation :  $x \longrightarrow y$ );  
Feedback exists between  $x$  and  $y$ : (Notation :  $x \longleftrightarrow y$ ).

#### 6.6.4 A Three-Variable Stock Market System

This section examines the relationships among the stock market, money supply and inflation. Prior research has shown that these three variables are linked. First, despite the Fisher effect, inflation has generally been shown to exhibit a negative relationship with the stock market [see Fama and Schwert (1977) and DeFina (1991)]. The reasons that have been advanced to explain the relationship include inflationary expectations, fixed price nominal contracts and the tax shield effects associated with depreciable fixed assets.<sup>25</sup> However as Stulz (1986) argues this relationship is dependent also on money growth.

Second, announcements of the money supply have been shown to convey a valuable information signal to the stock market. While there is some conjecture as to the sign of the relationship, it is generally accepted that a negative relationship exists between the money supply and stock returns. The general theory advanced is that the linkage between the money supply and interest rates affects economic activity and corporate

profits. However, there are questions over whether the real rate of interest is affected. Two main hypotheses<sup>26</sup> have emerged. First, changes in the money supply may alter expectations about monetary policy. An increase in the money supply may foreshadow a future tightening of monetary policy from the Central Bank resulting in expectations of higher interest rates, which in turn act to depress stock prices through both a rise in the real rate and a reduction in economic activity. Second, an increase in the money supply may raise expectations of higher inflation which in turn leads to higher interest rates through the inflation premium in nominal interest rates. As discussed above, higher expected inflation decreases stock prices. Both of these hypotheses suggest a negative sign on the relationship between money supply and the stock market which is generally supported by the evidence. Hardouvelis (1988) shows that increases in the money supply induce rises in interest rates. Moreover, Pearce and Roley (1985) and Jain (1988) find evidence of a significant negative relationship between unexpected money supply signals and stock market movements.

Finally, the third interaction in the system is the linkage between inflation and the money supply. This relationship is well-known and rooted in monetary theory [eg. Mishkin (1992)]. Despite arguments over the influence of lags and the money multiplier, the economic relationship is well established. Of note, the purpose here is not to test in detail hypotheses surrounding these variables, but rather to illustrate how relationships in the financial markets can be tested.

---

<sup>25</sup> DeFina (1991) provides a good overview of the various arguments.

<sup>26</sup> Another hypothesis, suggesting that higher money supply leads to increased general price levels, could be proposed. This will cause spillovers into company profitability, thereby increasing the stock prices. However in the current application this hypothesis is not supported by the empirical findings.

The following data are used in the test. The focus is the Australian market, both because of the ease of data availability and the lack of previous research in this area in the Australian market.<sup>27</sup> The All Ordinaries Index (AOI) is used as the stock market indicator. The AOI is a broad market indicator with coverage of around 320 stocks representing about 90-95% of total market capitalisation. The index is value-weighted and calculated on the basis of market capitalisation of the constituent stocks traded on the Australian Stock Exchange. Money supply is measured by the standard stock of money (M3).<sup>28</sup> Inflation is measured as the seasonally adjusted consumer price indices for Australia (CPI<sub>AUS</sub>). The CPI measures the aggregate price behaviour of all consumer goods and services and is commonly used by government and industry in Australia to adjust for cost-of-living allowances in wage and benefit contracts. Data are collected from DataStream™ over the period June 1981 through December 1999. While money supply and the stock market index are available over shorter frequencies, CPI figures are produced on a quarterly basis, and hence this forms the basis for the sampling frequency.

---

<sup>27</sup> The three variable system proposed here could be tested in any other market.

<sup>28</sup> M3 is a common measure of the money supply and is used in Reserve Bank targeting. While an alternative measure of M2 comprises money that can be spent immediately and assets invested for the short term, M3 consists of the sum of M2 plus large deposits. These deposits include institutional money-market funds and agreements among banks. Since M3 comprises M2, we employ M3 in the test.

The Dickey and Pantula (1987) procedure is used to test for the presence of more than one unit root. The procedure rejects the hypothesis of three unit roots for both  $\log(\text{CPI}_{\text{AUS}})$  and  $\log(\text{M3})$  at the 5 percent level, and the hypothesis of two unit roots for  $\log(\text{AOI})$ . Subsequently, the procedure accepts the hypothesis that both  $\log(\text{CPI}_{\text{AUS}})$  and  $\log(\text{M3})$  have two unit roots and  $\log(\text{AOI})$  has one unit root.

The procedure described in Section 6.6 is then utilised to identify the specification for the VECM formed by these variables. The results using SC are presented in Table 6.7. In addition, to check the adequacy of the model fit, the strategy suggested in Tiao and Tsay (1989) and Penm *et al* (1997) is used, with the proposed Penm and Terrell (1984) algorithm applied to test the residual vector series, using the SC criterion. The results in Table 6.7 support the residual vector being a white noise process.

The results are generally consistent with economic intuition and prior evidence. The causality identified in the selected ZNZ patterned VECM confirms that M3 is an independent source of financial and economic disturbance, and an indirect causality exists from M3 through  $\text{CPI}_{\text{AUS}}$  to AOI. This result supports the impact that money supply has on stock prices through inflationary pressures. Of note, the presence of this indirect causality cannot be detected from inspecting the nonzero elements in all their coefficient matrices of a conventional full-order VECM. Among  $\text{CPI}_{\text{AUS}}$ , M3 and AOI, two cointegrating vectors are found. The first selected cointegrating vector supports that  $\Delta \log(\text{AOI})$  is stationary. The second selected cointegrating vector confirms that  $\log(\text{CPI}_{\text{AUS}})$ ,  $\log(\text{AOI})$  and  $\Delta \log(\text{AOI})$

are cointegrated with  $\log(M3)$ . This indicates that both  $\log(CPI_{AUS})$  and  $\log(M3)$  are  $CI(2,1)$  processes, not  $CI(2,2)$  processes as described in Engle and Granger (1987). The positive sign between  $\log(M3)$  and  $\log(CPI_{AUS})$  and the negative sign between  $\log(M3)$  and  $\log(AOI)$  is consistent with the hypothesis discussed above that increases in the money supply leading to an increase in inflation, which in turn is taken by the stock market as a negative signal.

**Table 6.7**

**The VECM<sup>a</sup> for the relationship linking money supply, inflation and stock market indicator for Australia**

Variables:  $y_t^1 = \log(M3)$ ,  $y_t^2 = \log(CPI_{AUS})$ ,  $y_t^3 = \log(AOI)$ .

Sample Period: 1981.II to 1999.IV

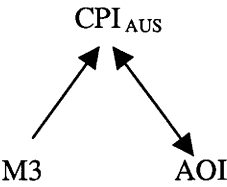
$  \begin{bmatrix} \Delta^2 y_t^1 \\ \Delta^2 y_t^2 \\ \Delta^2 y_t^3 \end{bmatrix} + \begin{bmatrix} 0.340 & 0.0 & 0.0 \\ (0.110) & & \\ 0.148 & 0.468 & 0.0 \\ (0.105) & (0.091) & \\ 0.0 & 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} \Delta^2 y_{t-1}^1 \\ \Delta^2 y_{t-1}^2 \\ \Delta^2 y_{t-1}^3 \end{bmatrix} + \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.155 & 0.0 \\ & (0.092) & \end{bmatrix} \begin{bmatrix} \Delta^2 y_{t-3}^1 \\ \Delta^2 y_{t-3}^2 \\ \Delta^2 y_{t-3}^3 \end{bmatrix} +  $									
$  \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ -0.045 & 0.388 & -0.105 \\ (0.020) & (0.105) & (0.070) \\ 0.0 & 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} y_{t-1}^1 \\ y_{t-1}^2 \\ y_{t-1}^3 \end{bmatrix} + \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.220 \\ 0.0 & 0.0 & 1.081 \\ & & (0.115) \end{bmatrix} \begin{bmatrix} \Delta y_{t-1}^1 \\ \Delta y_{t-1}^2 \\ \Delta y_{t-1}^3 \end{bmatrix} = \varepsilon(t)  $									
$  \hat{\alpha} = \begin{bmatrix} 0.0 & 0.0 \\ 0.0 & -0.045 \\ & (0.010) \\ 1.081 & 0.0 \\ (0.115) & \end{bmatrix} \hat{\beta} = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ & & & & & (0.106) \\ 1.0 & -8.433 & 2.278 & 0.0 & 0.0 & -4.789 \\ (0.452) & (2.285) & (1.532) & & & (2.660) \end{bmatrix}  $									
Residual analysis <sup>b</sup>	Existing lags	0	1	2	3	1,2	1,3	2,3	1,2,3

Value of SC	1.0	1.019	1.010	1.018	1.031	1.038	1.029	1.050
-------------	-----	-------	-------	-------	-------	-------	-------	-------

Long-term Cointegrating Relationship Identified:

- 1)  $\Delta \log(\text{AOI})$  is stationary
- 2)  $0.118 \log(\text{M3}) = \log(\text{CPI}_{\text{AUS}}) - 0.270 \log(\text{AOI}) + 0.568 \Delta \log(\text{AOI})$

Granger Causal Pattern<sup>c</sup> Recognised:



- a) Model selected by SC Using the GLS Procedure. Standard errors in parentheses.  $\Delta$  denotes first difference.
- b) For simplicity, the values of SC for  $q > 3$  are not presented, but can be supplied to readers upon request.
- c)  $x$  Granger-causes  $y$ : (Notation :  $x \longrightarrow y$ );  
Feedback exists between  $x$  and  $y$ : (Notation :  $x \longleftrightarrow y$ ).

### 6.7 Summary

This chapter consists of three parts. In the first part, the sensitivity of PPP testing to the nature of the unit root tests is demonstrated. Three unit root tests are employed for fourteen real bilateral exchange rates. The first two - the augmented Dickey-Fuller test and the Phillips-Perron test - are procedures using the null hypothesis of a unit root. The third, the Kwiatkowski *et al* (1992) test, is a procedure which employs the null hypothesis of stationarity. The results are contradictory. PPP is consistently accepted or rejected for only four exchange rates. For the remaining ten, PPP is accepted using the Kwiatkowski *et al* procedure, but rejected by the augmented Dickey-Fuller test and the Phillips-Perron test.

In the second part of this chapter, a new test procedure for PPP is introduced. In this procedure PPP is tested in the framework of subset VECM with zero coefficients. Since VECM modelling can accommodate both long-term and dynamic responses, this approach is different from those based on unit root tests. This new procedure adds to the available methods of testing for PPP.

The results of this procedure are promising. Both the necessary and the necessary and sufficient conditions for PPP are sequentially tested. Of the fourteen exchange rates investigated, the necessary condition is found acceptable for seven and the necessary and sufficient condition cannot be rejected for three. These results support the existence of PPP, at least for some exchange rates. In analysing these results, this thesis also finds that the short-term dynamic structure may be an important influence in testing the necessary and sufficient condition for PPP. This suggests that a more sophisticated specification could be useful in improving the test for PPP [see Penm *et al* (2001b)].

The third part of this chapter develops an effective algorithm to select the optimal ZNZ patterned cointegrating and loading vectors in a ZNZ patterned VECM for an  $I(2)$  system. Many financial series are of order  $I(2)$  and hence the procedure developed has substantial applicability. In addition, testing for PPP using an  $I(2)$  analysis has been conducted. The outcome confirms support for the necessary condition of the PPP hypothesis for the bilateral exchange rate between the Australian and US Dollar. The inter-relationships between the stock market, money supply and inflation are also studied and the results are generally consistent with both theory and prior evidence.



## CHAPTER 7

### NON-LINEAR FORECASTING USING NEURAL NETWORKS

#### 7.1 Introduction

Neural networks have been used in financial modelling and for simulation for a number of years [see Kim and Chan (1998), Refenes *et al* (1995)]. Zhang *et al* (1998) provide a survey of neural networks. These networks provide a procedure that generates *ex ante* forecasts as a non-linear function of historical observations. Chakraborty *et al* (1992) show that neural networks can be suitable for forecasting multivariate time-series, while Teräsvirta *et al* (1993) demonstrate the power of the neural network linearity test.

Conventionally, researchers have proposed procedures for conducting ‘full-order’ neural networks with fixed connections between nodes in all layers. However a full-order neural net has difficulty in practical modelling of subset time-series systems. Consequently this results in poor *ex ante* forecasting performance due to over-parameterised modelling. Hence, there is a need for relaxation of the assumption of fixed connections between nodes in all layers.

In this chapter an extension to the structure of neural networks is introduced to increase their modelling power for subset time-series systems. The extended neural networks are likely to improve the performance of financial modelling due to their

parsimonious model structure. At the same time they provide simulation instruments that can be used for analysing market properties.

This chapter also presents a numerically robust lattice-ladder learning algorithm that sequentially selects the best specification of a subset time-series system using extended neural networks. The algorithm is suitable for a situation in which the structure of the extended neural network is evolving. It is able to extend the relevance of multi-layered neural networks and so more effectively model a greater array of time-series applications. The approach recognises that many connections between nodes in layers are unnecessary and can be deleted. Inhibitor arcs - reflecting inhibitive synapses-are then introduced.

The algorithm allows for connections between nodes in different layers, which have variable strengths at different points of time, by introducing additional excitatory arcs - reflecting excitatory synapses. The ability to resolve both time and order updating leads to the optimal synaptic weight updating, which allows for the optimal dynamic node creation/deletion within the extended neural network. Thus the algorithm first uses the new available observation and regards the structure of the extended neural network as evolving, so the weights of the structure can be updated. Second, if the order lag structure is also evolving, the algorithm can utilise the order updating algorithm continuously to recheck and update the lag structure of the network.

The remaining sections are organised as follows: Section 7.2 introduces the extended neural network structure for subset time-series modelling. Section 7.3 demonstrates the overall lattice-ladder learning algorithm for extended neural networks. Section 7.4 provides an illustration, in which the proposed learning algorithm is used to describe the provision of all possible lattice-ladder structures for subset vector rational distributed lag (VRDL) modelling and all possible lattice structures for subset VAR modelling. Section 7.5 employs two case studies to demonstrate the usefulness of the algorithm in tests of causality between two equity market indicators and the inter-relationship in application to the stock market. In Section 7.6 a brief summary is provided to conclude the chapter.

## 7.2 The Extended Neural Network Structure for Subset Time-series Modelling

Consider a vector time-series model of the form

$$z(t) + \sum_{i=1}^p h'_i y(t+1-i) = \epsilon^h(t), \quad (7.1)$$

where  $h'_i$ ,  $i=1,2,\dots,p$  are  $g \times r$  parameter matrices,  $\epsilon^h(t)$  is a  $g \times 1$  stationary process with  $E\{\epsilon^h(t)\} = 0$  and

$$E\{\epsilon^h(t)\epsilon'^h(t-\tau)\} = \begin{cases} \Omega & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}.$$

Equation (7.1) and properties associated with  $\varepsilon^h(t)$  together constitute VRDL, which involves a  $g$ -dimensional regressand vector  $z(t)$  and an  $r$ -dimensional regressor vector  $y(t)$ .<sup>29</sup>

In subset VRDL modelling, the predictor of a subset VRDL system can be described as

$$\hat{z}(t) = -\sum_{i=1}^p h'_i(I_s)y(t+1-i), \quad (7.2)$$

where  $h'_i(I_s) = 0$  if  $i \in I_s$ . Thus  $I_s$  specifies the integers between 1 and  $p-1$  that correspond to excluded entries<sup>1</sup>. A VRDL model can serve as an infinite moving average representation of a rational vector AR, ARMA or ARMAX model [see Penm *et al* (1993) and (1999)]. The use of VRDL models in economic and financial time-series is versatile. Mittnik (1989) successfully applies the VRDL to time-series forecasting using balanced state space representations. Holmes and Hutter (1989) suggest the use of a subset VRDL system to assess the relationship between  $z(t)$  in (7.1) and the set of current and lagged  $y(t)$  where there is a continuous or a random delay.

---

<sup>29</sup> In practice the assumption that  $E\{\varepsilon^h(t)\} = 0$ , means that the vector data of both  $z(t)$  and  $y(t)$  need to be mean-corrected. For simplicity, mean-correction is ignored, as its presence would make no difference to any of the findings presented in this chapter.

Watanabe *et al* (1992) study a two-layered linear neural network as a physical structure of a full-order autoregressive moving-average model and find that no hidden layer suffices in the network for the learning requirements as  $y(t)$  comprises the first-order terms  $y_1(t), y_2(t), \dots, y_r(t)$ .

This chapter develops an extended two-layered neural network that can be easily applied to a time-series system with a subset structure. To achieve this objective two types of connection, namely inhibitor arc and switchable connection, are introduced to extend the neural network structure. The inhibitor arc was introduced to network theory by Petri (1979) and the associated connection strength for all these arcs is constrained to zero at all times. The switchable connection is obtained from switching theory and the strength is switchable between zero and non-zero at any time. The introduction of these connections increases the modelling and analytical power of the neural networks. This new feature supports neural nets as a tool for the study of subset time-series systems, which are common in financial markets.

As described in Chapter 1 traditional full-order time-series models, including VAR and VECM models, have become increasingly popular in the analysis of financial markets. Standard full-order time-series models assume nonzero entries in all their coefficient matrices. However applications of time-series models to financial market data have revealed that zero entries are indeed possible [see Brailsford *et al* (2001a)]. In such cases, the use of full-order models may lead to incorrect inferences and inferior forecasts. This chapter develops and investigates extended neural networks to select the model's optimal order and ZNZ pattern

determination. If key linear and non-linear interactions among variables are captured accurately, the chosen ZNZ system can improve the modelling and simulation performance.

If  $y(t)$  does not only include the first-order terms, but also contains the second-order terms  $y_1(t)y_2(t)$ ,  $y_1^2(t)$  and  $y_2^2(t)$ , then a polynomial neural network, with a single hidden layer investigated in Kwok and Yeung (1997), can be constructed as a physical structure of the subset VRDL model of (7.1). The hidden-node transfer function consists of a quadratic regression polynomial of two variables, and this network can approximate any non-linear function to a desired degree of accuracy. Section 3.7 of this thesis details this development.

It is often the case that a VRDL model, which works well in explaining the behaviour of a system covering a specific sample, may evolve over time because of political, economic, environmental or other external changes relating to financial time series. To incorporate such evolution, a learning algorithm is often included in a VRDL model so that the model can be updated over time. There are many well-developed computationally efficient and numerically robust recursive algorithms which can be employed to update the VRDL models [see Carayannis *et al* (1986)]. However, most of these algorithms are only applicable to full order models. In subset VRDL modelling, the commonly used learning algorithms for full order models are not applicable because the structure of the lag coefficients is estimated without the 'presence and absence' restrictions. As a result, it is necessary to develop a learning algorithm for subset VRDL models which include full-order models as a special case.

Section 7.3 focuses on a double (a-priori and a-posterior) lattice-ladder algorithm. The algorithm does not need to update the angle variable, but carries out in a more direct way the update of the optimal structure. This algorithm, which computes concurrently both the a-priori and a-posterior residuals in a recursion cycle, possesses better numerical accuracy and is less sensitive to roundoff errors than its direct matrix inversion counterpart. In addition, only the lattice algorithm is required for VAR modelling.

### 7.3 The Lattice-Ladder Algorithm for Subset VAR and VRDL Modelling

A double (a-priori/a-posterior) lattice-ladder recursive algorithm for subset VRDL and VAR modelling is introduced in this section. This is followed by providing the order-recursive algorithm, which can be used to initialise the lattice-ladder algorithm.

Let  $z(t)$  and  $y(t)$  be jointly stationary and zero mean vector time-series.<sup>30</sup> The subset VRDL(p) model can be described as:

$$z(t) + \sum_{i=1}^p h_i'(I_s)y(t+1-i) = \varepsilon^h(t). \quad (7.3)$$

As defined in (7.2),  $I_s$  represents an integer set with elements  $i_1, i_2, \dots, i_s$ ,  $1 \leq i_1 \leq \dots \leq i_s \leq p-1$ .  $h_i'(I_s) = 0$ , as  $i \in I_s$ . The disturbance variance-covariance

---

<sup>30</sup> A mean adjustment can be done first.

matrix is  $\Omega(I_s)$  and the covariance matrix, so-called partial correlation matrix between  $\varepsilon^h(t)$  and  $y(t-p)$ , is  $\Delta^h(I_s)$ . Given two finite data sample sets,  $\{y(1), \dots, y(T)\}$  and  $\{z(1), \dots, z(T)\}$ , and both  $y(j)=0$  and  $z(j)=0$  for  $j < 1$ , it is necessary to sequentially estimate all possible subset VRDL models from (7.3) using the prewindowed case investigated by Penm *et al* (1995). Since the actual scheme of (7.3) may not be order  $p$ , the resulting estimates of  $h_i$  is denoted by  $h_{p,T}(i)$ , where  $T$  is the sample size under examination. Then the predictor of (7.3) is of the form:

$$\hat{z}(i) = -H'_{p,T}(I_s)Y_{p,i}(I_s),$$

where  $Y_{p,i} = [y'(i) \ y'(i-1) \ \dots \ y'(i-p+1)]'$ ,

$$H_{p,T} = [h'_{p,T}(1) \ h'_{p,T}(2) \ \dots \ h'_{p,T}(p)]'.$$

$H_{p,T}(I_s)$  and  $Y_{p,i}(I_s)$  are formed by removing the  $(i_1), \dots, (i_s)$ 'th row block of  $H_{p,T}$  and  $Y_{p,i}$  respectively. The a-posterior prediction residual vector for observation  $i$  is defined as

$$\varepsilon^h_{p,i}(I_s) = z(i) + H'_{p,T}(I_s)Y_{p,i}(I_s), \quad (7.4)$$

and the a-priori prediction residual vector for observation  $i$  is

$$e^h_{p,i}(I_s) = z(i) + H'_{p,T-1}(I_s)Y_{p,i}(I_s). \quad (7.5)$$



In reality, many time-series systems present complex non-stationary features and cannot be modelled by assuming that  $y(t)$  and  $z(t)$  are stationary [see Bollerslev *et al* (1992)]. Thus, an estimate of the structure at time  $t$  should give a higher weight to the more recent observations and a lower weight to the observations of the more distant past. Thus, for a VRDL model fitted to these two sample sets using the prewindowed method of forming the sample covariance matrix, the following relationships can be established:

$$R_{p-1,T}(I_s)H_{p,T}(I_s) = -r_{p-1,T}(I_s), \text{ where } R_{p-1,T}(I_s) = \sum_{i=1}^T \lambda^{T-i} Y_{p,i}(I_s) Y'_{p,i}(I_s),$$

$$r_{p-1,T}(I_s) = \sum_{i=1}^T \lambda^{T-i} Y_{p,i}(I_s) z'(i), \text{ and } \Omega_{p,T}(I_s) = \sum_{i=1}^T \lambda^{T-i} \epsilon_{p,i}^h(I_s) \epsilon_{p,i}^{h'}(I_s),$$

where  $\lambda$ ,  $0 < \lambda \leq 1$ , is the fixed forgetting factor as described in Hannan and Deistler (1988).

The fixed forgetting factor is a data weighting process which gives more weight to recent observations and less weight to earlier data. The estimation method using the fixed forgetting factor can be considered as a type of kernel estimation (which will be discussed in Section 8.5). The use of the forgetting factor in statistical modelling has attracted significant attention in recent years. For example, Brailsford *et al* (2000) report the use of the forgetting factor in modelling and simulation of financial time-series, while Guo and Wu (1998) use the kernel regression to examine the exchange rate exposure of Taiwanese firms. The effect of their kernel is equivalent to the effect of a forgetting factor. Azimi-Sadjadi *et al* (1993) suggest the recursive updating procedure for the training process of a multi-

layer neural network, and Goto *et al* (1995) use the forgetting factor in the recursive least squares ladder algorithm for spectral estimation of a non-stationary process.

Next a further regressor  $y(t-p)$  is introduced into the VRDL model. For this enlarged VRDL( $p+1, I_s$ ) model, the following relationship is developed:

$$R_{p,T}(I_s) H_{p+1,T}(I_s) = - \begin{bmatrix} r_{p-1,T}(I_s) \\ v_{p,T}(I_s) \end{bmatrix},$$

$$\text{where } v_{p,T}(I_s) = \sum_{i=1}^T \lambda^{T-i} y(i-p) z'(i).$$

Following the work by Carayannis *et al* (1986), the forward subset VAR( $p$ ) model with the deleted elements  $i_1, i_2, \dots, i_s$  can be described as:

$$\sum_{i=0}^p a_i(I_s) y(t-i) = \varepsilon(t, I_s), \{ a_0(I_s) = I, a_i(I_s) = 0, \text{ as } i \in I_s, \} \quad (7.6)$$

where  $E\{\varepsilon(t, I_s)\} = 0$ ,  $E\{\varepsilon(t, I_s) y'(t-p-1)\} = \Delta(I_s)$  and  $E\{\varepsilon(t, I_s) \varepsilon'(t, I_s)\} = V(I_s)$  as  $k=0$ ;  $=0$  as  $k>0$ . In addition, a backward subset VAR( $p$ ) model [see Penm and Terrell (1982) and (1983)] can be considered as:

$$\sum_{i=0}^p b_i(M_s) y(t-p+i) = \bar{\varepsilon}(t, M_s), \{ b_0(M_s) = I, b_i(M_s) = 0, \text{ as } i \in M_s, \} \quad (7.7)$$

where  $M_s$  represents an integer set with elements  $m_1, m_2, \dots, m_s$ ,  $m_j = p+1-i_j$ ,  $j = 1, 2, \dots, s$ .

Suppose for a given sample  $\{y(1), \dots, y(T)\}$ , all possible subset VAR models from (7.6) and (7.7) are fitted. Since the actual scheme may not be of order  $p$ , the resulting estimates of  $a_k$  and  $b_k$  are denoted by  $a_{p,T}(k)$  and  $b_{p,T}(k)$  respectively. As a result:

$$A'_{p,T} = [a_{p,T}(1) \cdots a_{p,T}(p)], \quad B'_{p,T} = [b_{p,T}(p) \cdots b_{p,T}(1)].$$

For a forward VAR( $p, I_s$ ) fitted by this sample set, the following relationships are established:

$$R_{p,T}(I_s) = \sum_{i=1}^T \lambda^{T-i} Y_{p+1,i}(L_s) Y'_{p+1,i}(L_s), \text{ where } Y_{p+1,i}(L_s) = \begin{bmatrix} y(i) \\ Y_{p,i-1}(I_s) \end{bmatrix} = \begin{bmatrix} Y_{p,i}(L_s) \\ y(i-p) \end{bmatrix}, \quad (7.8)$$

$$R_{p,T}(I_s) \begin{bmatrix} I \\ A_{p,T}(I_s) \end{bmatrix} = \begin{bmatrix} V_{p,T}(I_s) \\ 0 \end{bmatrix}, \quad \epsilon_{p,i}(I_s) = y(i) + A'_{p,T}(I_s) Y_{p,i-1}(I_s),$$

where  $L_s$  represents an integer set with elements  $l_j, j=1, \dots, s$ , and  $l_j = i_j + 1$ .  $Y_{p+1,i}(L_s)$  and  $Y_{p,i}(L_s)$  are formed by removing the  $(l_1), \dots, (l_s)$ 'th row block of  $Y_{p+1,i}$  and  $Y_{p,i}$  respectively. Similarly  $A_{p,T}(I_s)$  and  $Y_{p,i-1}(I_s)$  are formed by removing the  $(i_1), \dots, (i_s)$ 'th row block of  $A_{p,T}$  and  $Y_{p,i-1}$  respectively.  $\epsilon_{p,i}(I_s)$  will be called the a-posterior forward AR residual vector to distinguish it from the a-prior forward AR residual vector defined as

$$e_{p,i}(I_s) = y(i) + A'_{p,T-1}(I_s) Y_{p,i-1}(I_s). \quad (7.9)$$

In addition, for the corresponding backward VAR( $p, M_s$ ), the following relationships are developed:

$$R_{p,T}(M_s) = \sum_{i=1}^T \lambda^{T-i} \begin{bmatrix} Y_{p,i}(M_s) \\ y(i-p) \end{bmatrix} \begin{bmatrix} Y'_{p,i}(M_s) & y'(i-p) \end{bmatrix} = \begin{bmatrix} D_{p,T}(M_s) & r^b_{p,T}(M_s) \\ r^{b'}_{p,T}(M_s) & v^b_{p,T}(M_s) \end{bmatrix},$$

$$R_{p,T}(M_s) \begin{bmatrix} B_{p,T}(M_s) \\ I \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{V}_{p,T}(M_s) \end{bmatrix}, \quad \bar{\epsilon}_{p,i}(M_s) = B'_{p,T}(M_s) Y_{p,i}(M_s) + y(i-p), \quad (7.10)$$

where  $B_{p,T}(M_s)$  and  $Y_{p,i}(M_s)$  are formed by removing the  $(p+1-m_1), \dots, (p+1-m_s)$ 'th row block of  $B_{p,T}$  and  $Y_{p,i}$  respectively. Thus  $R_{p,T}(M_s) = R_{p,T}(I_s)$ .

Next, the associated VAR and VRDL partial correlation matrices become:

$$\Delta_{p+1,T}(I_s) = v^b_{p+1,T}(M_s) + A'_{p,T}(I_s) r^b_{p,T-1}(M_s),$$

$$\text{and } \Delta^h_{p+1,T}(I_s) = v_{p,T}(I_s) + B'_{p,T}(M_s) r_{p-1,T}(I_s).$$

The forward and backward reflection matrices of the form become:

$$K^f_{p+1,T}(I_s) = -V^{-1}_{p,T}(I_s) \Delta'_{p+1,T}(I_s),$$

$$K^b_{p+1,T}(M_s) = -\bar{V}^{-1}_{p,T-1}(M_s) \Delta_{p+1,T}(I_s),$$

and the ladder gain matrix of the form becomes:

$$K_{p+1,T}(I_s) = -\bar{V}^{-1}_{p,T}(M_s) \Delta^h_{p+1,T}(I_s).$$

In addition,  $\bar{\epsilon}_{p,i}(\mathbf{M}_s)$  is called the a-posterior backward AR residual vector to distinguish it from the a-prior backward AR residual vector as defined by

$$\bar{\epsilon}_{p,i}(\mathbf{M}_s) = \begin{bmatrix} \mathbf{B}'_{p,T-1}(\mathbf{M}_s) & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{p,i}(\mathbf{M}_s) \\ y(i-p) \end{bmatrix}. \quad (7.11)$$

A double (a-priori/a-posterior) lattice recursive algorithm for subset VRDL and VAR modelling can then be developed as follows.

Lattice algorithm for subset VAR models

$$\mathbf{e}_{p+1,T}(\mathbf{I}_s) = \mathbf{e}_{p,T}(\mathbf{I}_s) + \mathbf{K}'_{p+1,T-1}(\mathbf{M}_s) \bar{\epsilon}_{p,T-1}(\mathbf{M}_s) \quad (7.12a)$$

$$\bar{\epsilon}_{p+1,T}(\mathbf{M}_s) = \bar{\epsilon}_{p,T-1}(\mathbf{M}_s) + \mathbf{K}'_{p+1,T-1}(\mathbf{I}_s) \mathbf{e}_{p,T}(\mathbf{I}_s) \quad (7.12b)$$

$$\Delta_{p+1,T}(\mathbf{I}_s) = \lambda \Delta_{p+1,T-1}(\mathbf{I}_s) + \bar{\epsilon}_{p,T-1}(\mathbf{M}_s) \mathbf{e}'_{p,T}(\mathbf{I}_s) \quad (7.12c)$$

$$\mathbf{K}^f_{p+1,T}(\mathbf{I}_s) = -\mathbf{V}_{p,T}^{-1}(\mathbf{I}_s) \Delta'_{p+1,T}(\mathbf{I}_s) \quad (7.12d)$$

$$\mathbf{K}^b_{p+1,T}(\mathbf{M}_s) = -\bar{\mathbf{V}}_{p,T-1}^{-1}(\mathbf{M}_s) \Delta_{p+1,T}(\mathbf{I}_s) \quad (7.12e)$$

$$\epsilon_{p+1,T}(\mathbf{I}_s) = \epsilon_{p,T}(\mathbf{I}_s) + \mathbf{K}'_{p+1,T}(\mathbf{M}_s) \bar{\epsilon}_{p,T-1}(\mathbf{M}_s) \quad (7.12f)$$

$$\bar{\epsilon}_{p+1,T}(\mathbf{M}_s) = \bar{\epsilon}_{p,T-1}(\mathbf{M}_s) + \mathbf{K}'_{p+1,T}(\mathbf{I}_s) \epsilon_{p,T}(\mathbf{I}_s) \quad (7.12g)$$

$$\bar{\mathbf{V}}_{p+1,T}(\mathbf{M}_s) = \lambda \bar{\mathbf{V}}_{p+1,T-1}(\mathbf{M}_s) + \bar{\epsilon}_{p+1,T}(\mathbf{M}_s) \bar{\epsilon}'_{p+1,T}(\mathbf{M}_s) \quad (7.12h)$$

$$\mathbf{V}_{p+1,T}(\mathbf{I}_s) = \lambda \mathbf{V}_{p+1,T-1}(\mathbf{I}_s) + \epsilon_{p+1,T}(\mathbf{I}_s) \epsilon'_{p+1,T}(\mathbf{I}_s) \quad (7.12i)$$

Ladder algorithm for subset VRDL models:

$$e_{p+1,T}^h(I_s) = e_{p,T}^h(I_s) + K'_{p+1,T-1}(I_s) \bar{e}_{p,T}(M_s) \quad (7.12j)$$

$$\Delta_{p+1,T}^h(I_s) = \lambda \Delta_{p+1,T-1}^h(I_s) + \bar{e}_{p,T}(M_s) e_{p,T}^h(I_s) \quad (7.12k)$$

$$K_{p+1,T}(I_s) = -\bar{V}_{p,T}^{-1}(M_s) \Delta_{p+1,T}^h(I_s) \quad (7.12l)$$

$$\varepsilon_{p+1,T}^h(I_s) = \varepsilon_{p,T}^h(I_s) + K'_{p+1,T}(I_s) \bar{e}_{p,T}(M_s) \quad (7.12m)$$

$$\Omega_{p+1,T}(I_s) = \lambda \Omega_{p+1,T-1}(I_s) + \varepsilon_{p+1,T}^h(I_s) e_{p+1,T}'^h(I_s). \quad (7.12n)$$

To initialise the lattice-ladder algorithm, the order update algorithm proposed by Penm and Terrell (1982) and (1983), is extended and incorporated in the current overall algorithm. If  $q$  denotes the number of non-zero coefficient matrices of the VAR model, the following order-update equations can be established.

Order updating for subset VAR models

$$\Delta = v_{p+1,T}^b(M_s) + A'_{p,T}(I_s) r_{p,T-1}^b(M_s)$$

$$K_{p+1,T}^f(I_s) = -V_{p,T}^{-1}(I_s) \Delta'$$

$$K_{p+1,T}^b(M_s) = -\bar{V}_{p,T-1}^{-1}(M_s) \Delta$$

$$A_{p+1,T}(I_s) = \begin{bmatrix} A_{p,T}(I_s) \\ 0 \end{bmatrix} + \begin{bmatrix} B_{p,T-1}(M_s) \\ I \end{bmatrix} K_{p+1,T}^b(M_s)$$

$$B_{p+1,T}(M_s) = \begin{bmatrix} 0 \\ B_{p,T-1}(M_s) \end{bmatrix} + \begin{bmatrix} I \\ A_{p,T}(I_s) \end{bmatrix} K_{p+1,T}^f(I_s)$$

$$\bar{V}_{p+1,T}(M_s) = \bar{V}_{p,T-1}(M_s) + \Delta K_{p+1,T}^f(I_s)$$

$$V_{p+1,T}(I_s) = V_{p,T}(I_s) + \Delta' K_{p+1,T}^b(M_s)$$

Order updating for subset VRDL models

$$\Delta_{p+1,T}^h(I_s) = v_{p,T}(I_s) + B'_{p,T}(M_s)r_{p-1,T}(I_s)$$

$$K_{p+1,T}(I_s) = -\bar{V}_{p,T}^{-1}(M_s)\Delta_{p+1,T}^h(I_s)$$

$$H_{p+1,T}(I_s) = \begin{bmatrix} H_{p,T}(I_s) \\ 0 \end{bmatrix} + \begin{bmatrix} B_{p,T}(M_s) \\ I \end{bmatrix} K_{p+1,T}(I_s)$$

$$\Omega_{p+1,T}(I_s) = \Omega_{p,T}(I_s) + \Delta_{p+1,T}^h(I_s)K_{p+1,T}(I_s).$$

In the special case, where the consecutive coefficient matrices  $a_{p-k}$  for the lags of  $y(t-p+k)$ ,  $k=0, 1, \dots, c$  ( $c < p-1$ ) of the forward VAR( $p$ ) model, fitted using the sample  $\{y(1), \dots, y(T)\}$ , are missing, the estimated coefficient matrices are null, i.e.  $a_{p-k,T}(I_s)=0$ . The corresponding coefficient matrices, a-prior and a-posterior forward prediction residual vectors, and  $V$  of the forward VAR( $p-c-1$ ) model are sufficient to continue the recursive estimations.

Similarly, if the consecutive coefficient matrices  $h_{p-k}$  of the VRDL( $p$ ) model are missing, the analogous relations will hold.

Again, in the special case where the consecutive coefficient matrices  $b_{p-k}$ ,  $k=0, 1, \dots, d$  ( $d < p-1$ ) of the backward VAR( $p$ ) model, fitted using the sample  $\{y(1), \dots, y(T)\}$ , are missing, the coefficient matrices, a-prior and a-posterior backward prediction residual vectors, and  $\bar{V}$  from the backward VAR( $p-d-1$ ) fitted using the sample  $\{y(1), \dots, y(T-d-1)\}$  allow the proposed lattice-ladder and order-update recursions.

The order update algorithm is well suited to the initial stage of the lattice-ladder recursive algorithm. After the initialisation is carried out and all necessary lattice-ladder parameter matrices at  $t=T$  are obtained, the lattice-ladder recursive algorithm can be carried out from  $T$  to  $T+1$ ,  $T+2$ , and so on. A model information criterion is then used to select the lag structure of the optimal subset VRDL and that of the optimal subset VAR at each time instant. Finally, the associated coefficient matrices  $H_i(I_s)$  and  $A_i(I_s)$  can be estimated. A detailed illustration for  $K=Q=4$  is provided below in Section 7.4.

Of note, as mentioned in Penm *et al* (1995), recursively updated algorithms for subset time-series models are often desirable, when observation measurements exhibit some form of periodic behaviour when covering a range of different measurement periods [see Penm (1977)]. For instance financial data can be measured in periods of seconds, hours, and days, and from time-series systems where periodic variation exists.



## 7.4 Illustration

Suppose there is a fit of all possible subset VAR models for  $y(t) = \{y_1, y_2, \dots, y_{81}, \dots\}$  up to and including the maximum lag  $K=4$ , and all possible subset VRDL systems,  $z(t) = \{z_1, z_2, \dots, z_{81}, \dots\}$  and  $y(t) = \{y_1, y_2, \dots, y_{81}, \dots\}$ , up to and including the maximum lag  $Q=K$  at  $T=79, 80, 81$  and so on. The following description would present a suitable method. Note that numerals 0, 1, 2 represent particular lags for  $y(t-s)$ ,  $s \geq 0$ , underlined numerals 1, 2,... represent particular leads for  $y(t+s)$ ,  $s \geq 1$ , in VAR models, and italic numerals *0*, *1*, *2* represent particular lags for  $y(t-s)$ ,  $s \geq 0$  in VRDL systems.

In the initial stage, the parameter matrices  $A_{p,T}$ ,  $V_{p,T}$ ,  $B_{p,T}$ ,  $\bar{V}_{p,T}$  for each necessary subset VAR model, and  $H_{p,T}$  and  $\Omega_{p,T}$  for each necessary subset VRDL model, can be obtained using the order update recursions at  $T=76, 77, 78$  and  $79$ .<sup>31</sup> All necessary models for performing the order update recursions are listed in Table 7.1. To achieve this end,  $K$  should be set at 0 in the first recursion for each subset model at  $T=76$  (i.e.  $79-4+1$ ), be set at 1 in the second recursion at  $T=77$ , and then increase linearly in time until it reaches 4 at  $T=79$ . Next, the lattice-ladder recursions are conducted from  $T=79$  to 80, 81, and so on. A procedure for performing the order update recursions is as follows:

---

<sup>31</sup> In order to conduct the lattice-ladder recursions from  $T=79$ , the parameter matrices of some models at  $T=76, 77$  and  $78$  are required (see Table 7.1). Use is made of the order update recursions to estimate these models.

(1) All one-lag forward and backward VAR models (1,1), (2,2), (3,3), and (4,4) at  $T=76, 77, 78$  and  $79$  and all one-lag VRDL models  $0, 1, 2$  at  $T=79$  must be fitted to start the recursions.

(2) Carry out the order update recursions to fit all 2-lag forward and backward VAR models at  $T=77, 78, 79$ . All estimates of each 2-lag VRDL system at  $T=79$  can also be obtained. Stage 1 in Table 7.1 presents the required models for performing the recursions.

(3) Again, carry out the order update recursions to estimate all 3-lag subset VAR models at  $T=78$  and  $79$  and VRDL systems at  $T=79$ . Stage 2 in Table 7.1 illustrates this situation.

(4) Obtain the estimates of the full order forward and backward VAR models with  $K=4$  and the full order VRDL with  $Q=K$  at  $T=79$  using the order update recursions. Stage 3 in Table 7.1 lists all models required.

To this stage, both  $\Delta$  for each subset VAR and  $\Delta^h$  for each subset VRDL are obtained.

Assuming that all relevant quantities up to and including time  $T$  are known, the procedure for carrying out the lattice-ladder recursions from  $T$  to  $T+1$  is as follows:

*Step 1:* To begin with, all 1-lag forward and backward VAR models and all 1-lag VRDL models must be fitted. The associated prediction residual vectors are then computed:

$$e_{p,T}^h(I_s), \bar{e}_{p,T}^h(I_s), e_{p,T}(I_s), \bar{e}_{p,T}(I_s), e_{p,T}(M_s), \bar{e}_{p,T}(M_s)$$

by using the equations (7.5), (7.4), (7.9), (7.6), (7.11) and (7.7) respectively.

*Step 2:* For all 2-lag models,  $e_{p+1,T+1}(I_s)$ ,  $\bar{e}_{p+1,T+1}(M_s)$ ,  $e_{p+1,T+1}^h(I_s)$ ,  $p=1$  can be obtained using (7.12a), (7.12b) and (7.12j).

*Step 3:* The partial correlation, reflection coefficient and ladder gain matrices can be updated using (7.12c), (7.12d), (7.12e), (7.12k) and (7.12l).

*Step 4:* (7.12f), (7.12g) and (7.12m) are then used to estimate  $\epsilon_{p+1,T+1}(I_s)$ ,  $\bar{\epsilon}_{p+1,T+1}(M_s)$ ,  $\epsilon_{p+1,T+1}^h(I_s)$ ,  $p=1$ .

*Step 5:* The residual variance-covariance matrices can be estimated using (7.12h), (7.12i) and (7.12n).

*Step 6:* Repeat Steps 2 to 5 for all 3-lag models, and then all 4-lag models.

Furthermore, an order selection criterion for the vector case as suggested by Hannan and Deistler (1988), could be modified to use at each time instant with the proposed method to select the optimal subset VAR and VRDL models. From now on, MHQC is used as an abbreviation of the modified criterion, which is defined by

$$\text{MHQC} = \log|\hat{G}| + [2\log\log f(T)/f(T)]S, \quad f(T) = \sum_{i=1}^T \lambda^{T-i},$$

where  $f(T)$  is the effective sample size,  $S$  is the number of functionally independent parameters, and  $\hat{G}$  is the estimate of  $V_{p,T}(I_s)$  for the VAR and the estimate of  $\Omega_{p,T}(I_s)$  for the VRDL.

**Table 7.1**  
**The necessary VRDLs and the forward and backward VARs<sup>a</sup>**

the (p+1)th model		the required pth models	
T		T	T
<b>Stage 1</b>			
12	(79)	1	(79) <u>1</u> (78)
13	(79)	1	(79) <u>2</u> (78)
14	(79)	1	(79) <u>3</u> (78)
23	(79)	2	(79) <u>1</u> (77)
24	(79)	2	(79) <u>2</u> (77)
34	(79)	3	(79) <u>1</u> (76)
<u>12</u>	(79)	1	(79) <u>1</u> (78)
<u>13</u>	(79)	2	(79) <u>1</u> (77)
<u>14</u>	(79)	3	(79) <u>1</u> (76)
<u>23</u>	(79)	1	(79) <u>2</u> (78)
<u>24</u>	(79)	2	(79) <u>2</u> (77)
<u>34</u>	(79)	1	(79) <u>3</u> (78)
<i>01</i>	(79)	<i>0</i>	(79) <u>1</u> (79)
<i>02</i>	(79)	<i>0</i>	(79) <u>2</u> (79)
<i>03</i>	(79)	<i>0</i>	(79) <u>3</u> (79)
<i>12</i>	(79)	<i>1</i>	(79) <u>1</u> (78)
<i>13</i>	(79)	<i>1</i>	(79) <u>2</u> (78)
<i>23</i>	(79)	<i>2</i>	(79) <u>1</u> (77)
<b>Stage 2</b>			
123	(79)	12	(79) <u>12</u> (78)
124	(79)	12	(79) <u>23</u> (78)
134	(79)	13	(79) <u>13</u> (78)
234	(79)	23	(79) <u>12</u> (77)
<u>123</u>	(79)	12	(79) <u>12</u> (78)
<u>124</u>	(79)	23	(79) <u>12</u> (77)
<u>134</u>	(79)	13	(79) <u>13</u> (78)
<u>234</u>	(79)	12	(79) <u>23</u> (78)
<i>012</i>	(79)	<i>01</i>	(79) <u>12</u> (79)
<i>013</i>	(79)	<i>01</i>	(79) <u>23</u> (79)
<i>023</i>	(79)	<i>02</i>	(79) <u>13</u> (79)
<i>123</i>	(79)	<i>12</i>	(79) <u>12</u> (78)
<b>Stage 3</b>			
1234	(79)	123	(79) <u>123</u> (78)
<u>1234</u>	(79)	123	(79) <u>123</u> (78)
<i>0123</i>	(79)	<i>012</i>	(79) <u>123</u> (79)

<sup>a</sup>These are required to estimate all 2,3, or 4-lag subsets at T=79, using the order update recursions, and to update their a-priori/a-posterior errors, using the double (a-priori/a-posterior) lattice-ladder structure.

Possible zero constraints should be considered in the non-zero coefficient matrices of the optimal subset VAR and VRDL models. Geweke and Singleton (1981) introduced prior zero constraints in the coefficient matrices of a VRDL model to test hypotheses on the relations between observed economic time-series and latent factors. Previous studies [see Penm and Terrell (1984a) and Penm *et al* (1992)] show that the proposed search algorithm can, with minor modification, be adapted to select the optimum subset VAR and VRDL models with zero constraints using the prewindowed case.

## **7.5 Empirical Examples**

This section presents two applications to illustrate the practical use of the algorithm. The first application concerns a causality relationship between the stock and futures markets. The second application concerns the relationship between prices of an individual share and the underlying stock market.

### **7.5.1 The Causal Relationship between AOI and SPI**

The extended two-layered neural network described in Section 3.7 of this thesis can be used to model the relationship between financial time-series. In this example, the relationship between the Australian stock and futures markets is used. A futures contract is one of the most important hedging instruments for the underlying asset. Stock index futures have many attractive hedging benefits for a trader who wishes to trade the underlying stock portfolio corresponding to the

index. As described in Section 6.6.4, in Australia the main stock market indicator is the All Ordinaries Index (AOI). The index is calculated on the basis of market capitalisation of the constituent stocks traded on the Australian Stock Exchange. The Sydney Futures Exchange offers a futures contract on the AOI. This contract is available on a quarterly expiry date and is known as the Share Price Index (SPI) Futures Contract.

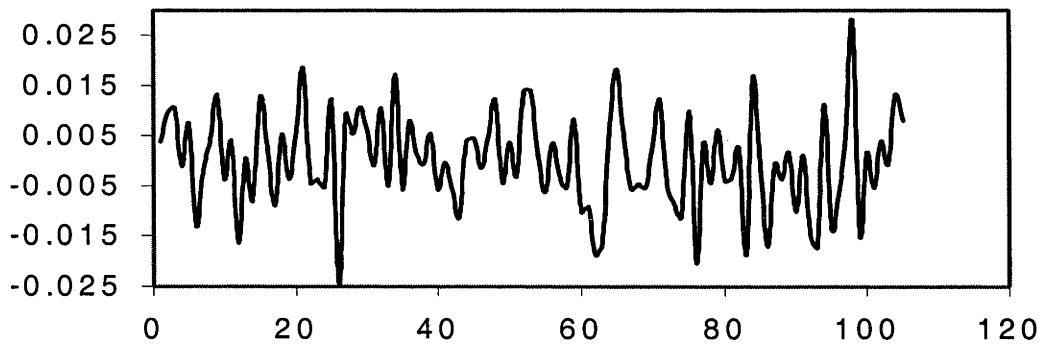
There is already a considerable literature examining the relationship between stock and futures market prices. The literature has either examined theoretical relationships between the markets through models such as the cost-of-carry [see Chung (1991), MacKinlay and Ramaswamy (1988), Brailsford and Hodgson (1997)], or examined the causality between the markets through lead-lag relationships, cointegration tests or bivariate spillover models [see Chan (1992), Chan *et al* (1991), Martens *et al* (1996)]. The general findings confirm a strong causality between the markets [see Wahab and Lashgari (1993), Abhyankar (1995)]. This relationship is not unexpected given the pricing relationship between the markets and the fact that the basis reduces to zero at the maturity of the futures contract. However there has been debate about the direction of causality, with the evidence generally indicating that the futures market leads the stock market. In particular, Chan (1992) has examined the lead-lag relation between returns of the Major Market cash index and returns of the Major Market Index futures and S&P 500 futures. His results indicate that the futures price is a leading indicator for the spot, when stock prices move together under market-wide movements. Tse (1995) has studied the causal relation between stock index futures and cash index prices in Japan, and documents that futures prices cause cash index prices.

Data on the AOI and SPI are sampled daily between 27 January and 29 June 1998. The AOI data are observed as the daily market closing index value whereas the SPI data are observed as the last traded price on each day in the June 98 contract. Graphs of log AOI and log SPI in first differences are shown in Figures 7.1 and 7.2. To test for the unit-roots for each plotted series, Microfit 4.0 is used to carry out the augmented Dickey-Fuller (ADF) unit root test. The 95 percent critical values for each test computed using the response surface estimates indicate that both log AOI and log SPI are non-stationary but both  $\Delta \log \text{AOI}$  and  $\Delta \log \text{SPI}$  are stationary for the period 27 January 1998 to 17 June 1998 ( $T=98$ ).  $\Delta \log \text{AOI}$  and  $\Delta \log \text{SPI}$  continue to be stationary for the extended period from 27 January 1998 to 29 June ( $T=106$ ), where  $T$  is the sample size now under consideration.



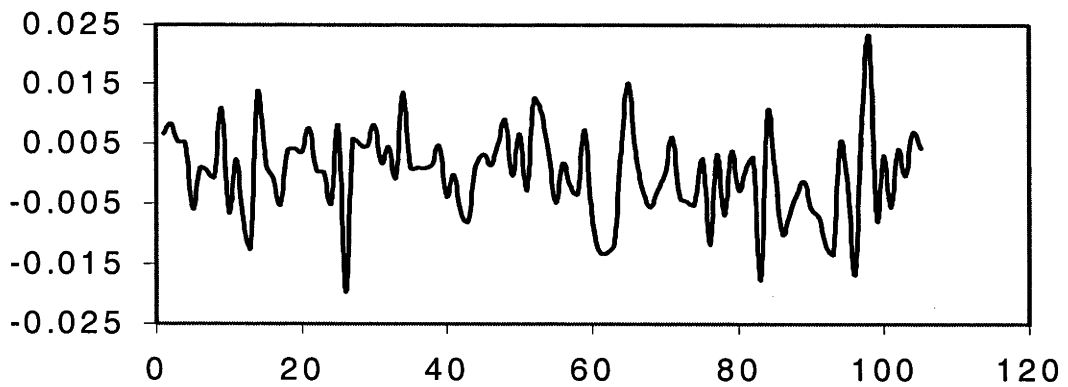
**Figure 7.1**

**Log AOI in first differences, daily: 27 January 1998 to 29 June 1998**



**Figure 7.2**

**Log SPI in first differences, daily: 27 January 1998 to 29 June 1998**



The extended two-layered neural network described in Section 3.7 is used to model the relationships between the AOI and SPI. In detecting the causal relationship from log SPI to log AOI, the variables used are  $z(t) = \log \text{AOI}$  and  $y(t) = \log \text{SPI}$ . As discussed above, neither log AOI nor log SPI are stationary. Therefore, the forgetting factor,  $\lambda$ , is incorporated to allow for the presence of non-stationarity. To begin, it is assumed  $Q=16$ , which corresponds to a three-week period (i.e. 15 business days). The order update and the lattice-ladder recursions described above are then used to select the 'optimal' specification of the distributed lag models at  $T=98, 99, \dots, 106$ .

For the causal relationship from log SPI to log AOI, the optimal distributed lag models with  $\lambda=0.99$  and  $0.985$  are presented in Table 7.2. The lower value of  $\lambda$  is consistent with strong persistence in market price fluctuations. For brevity, only the results obtained by the MHQC are presented. For cases where  $\lambda$  is less than  $0.985$ , the selected models are not reported due to the small effective sample size ( $<50$ ).

To assess the causality from log SPI to log AOI, a subset model for  $\lambda = 0.99$  with lags  $(0,1,10)$  was selected by the MHQC at  $T=98,99,\dots,102$ . A two-layered neural network can be constructed for this model. At  $T=103$ , the lag structure selected changes to  $(0,10)$ . The connection strength from the input unit representing  $y(t-1)$  to the output unit has switched from non-zero to zero. In this case, the predicted output, log AOI, is related to the current and previous inputs of log SPI. Also, the lag 0 indicates instantaneous causality between AOI and SPI. These results

indicate that both instantaneous and direct causal relationships exist from the futures market to the stock market even when emphasis is placed on recent data.

**Table 7.2**

**The VRDLs selected by MHQC for detecting the causal relationship from SPI to AOI**

Sample size (T)	Non-zero lag structure for $z(t)=\log \text{AOI}$ and $y(t)=\log \text{SPI}$		Pattern of causality <sup>a</sup>
	$\lambda=0.99$	$\lambda=0.985$	
98,99,100,101,102	0 1 10	0 10	$\log \text{SPI} \rightarrow \log \text{AOI}$
103,104,105,106	0 10	0 10	$\log \text{SPI} \rightarrow \log \text{AOI}$

(a)  $y \rightarrow z$ :  $y$  causes  $z$  directly and instantaneously.

For the causal relationship from  $\log \text{AOI}$  to  $\log \text{SPI}$ , Table 7.3 shows the optimal distributed lag models with  $\lambda=0.99$  and  $0.985$ . These results strongly support the existence of instantaneous causal and bidirectional feedback relationships between the Australian stock and index futures markets. These conclusions are generally consistent with those reported elsewhere in similar markets by Chan *et al* (1991), Chan (1992), Wahab and Lashgari (1993), and Abhyankar (1995).

In general these outcomes can be explained by reference to transaction costs, time delays in computing the index, execution costs, and measurement errors [see Chan (1992)]. In addition to speculators, some investors, particularly institutional investors, participate in the futures market for hedging purposes. Usually they take opposite positions in the stock market and the futures market at the same time, in order to hedge their exposure. Since they participate in both markets, price information will flow between the two markets. Therefore the finding of

instantaneous causal and bidirectional feedback relationships between these two markets is consistent with our prior hypotheses.

**Table 7.3**  
**The VRDLs selected by MHQC for detecting the causal relationship from**  
**AOI to SPI**

Sample size (T)	Non-zero Lag Structure for $z(t)=\log \text{ SPI}$ and $y(t)=\log \text{ AOI}$		Pattern of causality <sup>a</sup>
	$\lambda = 0.99$	$\lambda = 0.985$	
98,99,100,101,102	0 1 10	0 1 10	$\log \text{ AOI} \Rightarrow \log \text{ SPI}$
103,104,105,106	0 10	0 10	$\log \text{ AOI} \Rightarrow \log \text{ SPI}$

(a)  $y \Rightarrow z$ : y causes z directly and instantaneously.

Note there are other financial variables which could play a significant role in the stock-futures market analysis [see Brailsford and Hodgson (1997)]. This application merely demonstrates the usefulness of the extended, two-layered, neural network structure and the learning algorithm in time-series analysis

### 7.5.2 The Relationship between the Share Price of Telstra and AOI

The second application examines the non-linearities in the relationship between share prices and the underlying stock market. It focuses on the share price for Telstra Corporation Limited and the AOI. Telstra is one of Australia's largest companies. The traded shares represent the partially privatised Commonwealth Government's telecommunications organisation. The data used are daily series over the period 24 November 1997 to 30 March 1998 and were obtained from

DataStream™.<sup>32</sup> The value of Telstra's beta, 1.019, is calculated by regressing the Telstra's share returns against the returns of the stock market, where the AOI is used as a proxy for the market portfolio. This beta value reflects a strong relationship between the share price of Telstra and AOI.

The graph of the log of Telstra's share price is shown in Figure 7.3. The 95 percent critical values for the ADF unit root test indicate that the log of Telstra's share price is non-stationary. To demonstrate the usefulness of the proposed algorithm in a small sample environment, forecasting for period (T+1) is carried out by building subset VAR systems on the first T periods, using three-layer neural networks. The logarithms of the data are detrended and mean-corrected, rather than differenced for illustration purposes. Exponential forgetting was used with a forgetting factor 0.99. The optimal subset VAR models with zero constraints for T=75 through T=87 were identified, where  $y_1(t)$  = log of Telstra's share price,  $y_2(t)$  = log(AOI),  $y_3(t) = y_1(t)y_2(t)$ ,  $y_4(t) = y_1^2(t)$  and  $y_5(t) = y_2^2(t)$ . The specifications of all identified optimal VARs as a basis for constructing three-layer neural network are outlined in Table C.1 of Appendix C. One-step ahead forecasts based on each optimal VAR are calculated. For brevity, the forecasts are summarised in Table 7.4. The forecasts based on the naive random-walk model and the full-order VAR model are also shown for comparison purposes.

---

<sup>32</sup> Since then, on 31 March 1998, the Telstra Entity paid a fully franked interim dividend of seven cents per share. The 30th of March is chosen as the end period.

Figure 7.3

Log (Telstra's share price), daily: 24 November 1997 to 30 March 1998

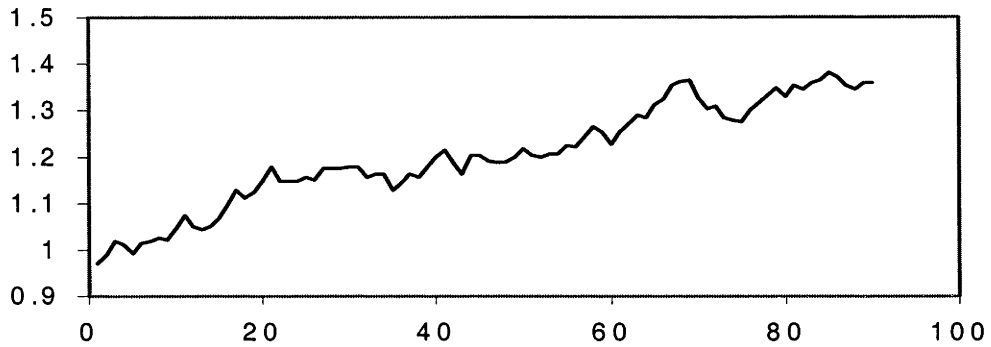


Table 7.4

ZNZ VAR, full-order VAR, and random-walk forecasts: Telstra share price

Trading days in 1998	T	Telstra (AUD\$)		ZNZ VAR			Full-order VAR			Random Walk	
		Share Price	Daily Price Change	Forecast Value	Error	Change Predicted	Forecast Value	Error	Change Predicted	Forecast Value	Error
10-03	77	3.67	+0.09	3.64	-0.03	+0.06	3.64	-0.03	+0.06	3.58	-0.09
11-03	78	3.73	+0.06	3.73	+0.00	+0.06	3.72	-0.01	+0.05	3.67	-0.06
12-03	79	3.79	+0.06	3.76	-0.03	+0.03	3.76	-0.03	+0.03	3.73	-0.06
13-03	80	3.85	+0.06	3.82	-0.03	+0.03	3.80	-0.05	+0.01	3.79	-0.06
16-03	81	3.78	-0.07	3.82	0.04	-0.03	3.84	0.06	-0.01	3.85	0.07
17-03	82	3.87	+0.09	3.86	-0.01	+0.08	3.80	-0.07	+0.02	3.78	-0.09
18-03	83	3.84	-0.03	3.83	-0.01	-0.04	3.88	0.04	+0.01	3.87	0.03
19-03	84	3.89	+0.05	3.86	-0.03	+0.02	3.86	-0.03	+0.02	3.84	-0.05
20-03	85	3.91	+0.02	3.90	-0.01	+0.01	3.87	-0.04	-0.02	3.89	-0.02
23-03	86	3.98	+0.07	3.94	-0.04	+0.03	3.90	-0.08	-0.01	3.91	-0.07
24-03	87	3.94	-0.04	3.96	0.02	-0.02	3.97	0.03	-0.01	3.98	0.04
25-03	88	3.87	-0.07	3.93	0.06	-0.01	3.97	0.10	+0.03	3.94	0.07
26-03	89	3.84	-0.03	3.87	0.03	+0.00	3.92	0.08	+0.05	3.87	0.03
27-03	90	3.89	+0.05	3.87	-0.02	+0.03	3.91	0.02	+0.07	3.84	-0.05
30-03	91	3.89	+0.00	3.89	+0.00	+0.00	3.90	0.01	+0.01	3.89	+0.00
RMSE <sup>a</sup>					0.0287			0.0524			0.0580

<sup>a</sup> RMSE =  $\sqrt{\frac{1}{15} \sum_{i=1}^{15} (\hat{y}_i - y_i)^2}$ , where  $\hat{y}_i$  and  $y_i$  denote the forecast value and the actual share price respectively.

It is commonly agreed that accurate forecasting in share markets is a difficult task. The approach adopted here focuses on models which can correctly simulate share price rises, falls or no change. The trading profits gained from the outcome of these models can still be offset by stock market frictions, such as stamp duty costs, capital gains taxes, and broker' fees. To examine the results presented in Table 7.4, the optimal ZNZ patterned VARs appear to perform well. Of the fifteen signs (when there were ten '+'s and five '-'s) of daily price change over the test period, ZNZ VARs successfully predicted fourteen of these, whereas full-order VAR effectively simulated ten and the random-walk model correctly forecast only one.

Theil's (1966) inequality coefficient,  $\sqrt{\frac{\sum (P_i - A_i)^2}{\sum A_i^2}}$ , is 0.0075 for the ZNZ VAR forecast, 0.0136 for the full-order VAR forecast, and 0.0151 for the random-walk model forecast, where  $(P_i, A_i)$  stands for a pair of predicted and observed changes. The root mean squared error (RMSE) error of the ZNZ VAR forecast is 54.86% of the prediction error of the full-order VAR forecast, and 49.55% of the random-walk forecast, respectively.

Of course many other non-parametric regression techniques and dynamic forgetting factor methods could play a significant role in simulations. The interest of this study is mainly to investigate how the algorithm may be utilised to improve the accuracy of financial simulations of share prices by using polynomial neural networks. Possible algorithms to improve the accuracy of the forecast magnitude are being investigated by various researchers.

## 7.6 Summary

In this chapter, a numerically robust lattice-ladder learning algorithm has been developed to sequentially select the best specification for a subset time-series system using neural networks. An extended two-layered neural net is developed to model the VRDL system with a subset structure. As the neural input vector,  $y(t)$ , includes second-order terms, the neural net with slight modification is extended to the cases of three-layers. The proposed construction method is simple to use and can be applied to an  $M$ -layered polynomial neural network with hidden layer nodes in layer  $m \in [1, M-2]$ . The overall lattice-ladder learning algorithm for extended neural networks avoids cumbersome matrix inversion and results in better numerical accuracy. Section 7.5 employs two case studies to demonstrate the usefulness of the algorithm. The first application investigates a causality relationship between the stock and futures markets. The second application examines the relationship between prices of an individual share and the underlying stock market. These two applications demonstrate the effectiveness of the proposed algorithms and widen the possible use of the forgetting factor in financial simulations and/or forecasting.

The following chapter introduces and develops the forgetting factor method in financial time-series analysis. Two procedures are proposed to select the forgetting factor in subset AR modelling. The first procedure uses the bootstrap to determine the value of a fixed forgetting factor. The second procedure starts from this base and applies the time-recursive maximum likelihood estimation for a dynamic forgetting factor. Two illustrations are presented to demonstrate the



usefulness of the proposed procedures. Improved forecasting performance arising from these proposed procedures is consistently achieved.

## CHAPTER 8

### SELECTING THE FORGETTING FACTOR IN SUBSET AUTOREGRESSIVE MODELLING

#### 8.1 Introduction

As noted in the previous chapter, the use of the forgetting factor in time series modelling has attracted significant attention in recent years. For example, in the recursive estimation of an autoregressive (AR) model, Hannan and Deistler (1988) apply the forgetting factor in model order determination for a non-stationary process. Goto *et al* (1995) use the forgetting factor in the recursive least squares ladder algorithm for spectral estimation of a non-stationary process. Azimi-Sadjadi *et al* (1993) incorporate the forgetting factor into a recursive weight updating procedure for the training process of a multilayer neural network.<sup>33</sup>

The forgetting factor is a data weighting process which gives more weight to recent observations and less weight to earlier data. Incorporating the forgetting factor in a time series model means that estimates of a model at time  $t$  give more weight to recent observations and less to past observations.

---

<sup>33</sup> Other studies which utilise the forgetting factor in time series analysis include Cho *et al* (1991) and Moscinski and Ogonowski (1995).

Traditionally the forgetting factor has been used in slowly time-varying linear models. A linear time-series model, which works well in explaining the behaviour of a process over a specific sample, may evolve slowly over time due to economic, political or structural changes. While a non-stationary series under such evolution may still be described by linear models, the forecasts obtained by allocating greater weight to more recent observations and ‘forgetting’ some of the past are likely to outperform alternatives in which such an allocation is not adopted. Hence, forecasting performance can be improved.

Consider an AR (p) model of the following form:

$$y(t) + \sum_{i=1}^p a_i y(t-i) = \varepsilon(t), \quad (8.1)$$

where  $\varepsilon(t)$  is an independent and identically distributed random process with  $E\{\varepsilon(t)\} = 0$  and  $E\{\varepsilon(t)\varepsilon(s)\} = \delta_{ts}\sigma^2$ , and the observations  $y(t)$   $\{t=1,2, \dots, T\}$  are available.

Following the approach of Hannan and Deistler (1988) a practical strategy for determining the value of the forgetting factor  $\lambda(t)$  is proposed as follows:

$$\lambda(t) = \begin{cases} \lambda & \text{if } 1 \leq t \leq S \\ 1 & \text{if } S < t \leq T \end{cases},$$

where ‘forgetting’ of the past occurs from time  $S$  to time  $1$ , and then no forgetting is involved from time  $S+1$  to time  $T$ . If  $\lambda=1$  for every  $t$ , then the ordinary least squares solution is obtained, but if  $0 < \lambda < 1$  is chosen, the past is weighted down

geometrically from time  $S$ . This means that the coefficients in (8.1) are estimated to minimise:

$$\sum_{t=1}^T \kappa_{T-t} [y(t) - \sum_{i=1}^p a_i y(t-i)]^2,$$

where the forgetting profile  $\kappa_{T-t}$ , which increases with  $t$  from time 1 to time  $S$  for a given  $S$ , is defined as:

$$\kappa_{T-t} = \begin{cases} \lambda^{S-t+1} & 1 \leq t \leq S \\ 1 & S < t \leq T \end{cases}. \quad (8.2)$$

For convenience, it is assumed that ‘forgetting’ of the past begins from time  $T-1$  immediately, that is  $S = T-1$ .

One important issue relating to the use of a forgetting factor in AR modelling is how to determine its value in various applications. The conventional methods are mostly based on arbitrary or personal choices. In this study two procedures are proposed to determine the value of the forgetting factor in subset AR modelling. In the first procedure the bootstrap is used to determine the value of a forgetting factor which will be fixed over a given sample (the fixed forgetting factor). The second procedure is based on the time-update recursive maximum likelihood (TRML) estimation for on-line time updating [see Hannan and Deistler (1988)]. This procedure presents a computationally efficient method for the analysis of time series data, i.e. the parameters at the  $k$ th stage may be explicitly estimated from the required information available from the  $(k-1)$ th stage for  $k=2,3,4,\dots$ . Subsequently

the value of a forgetting factor which is allowed to change can be determined as the time updating process is conducted (the dynamic forgetting factor).

Furthermore, the order selection criterion, that is MHQC as suggested in Section 7.4, could be used to select the optimal subset AR model. MHQC is defined by

$$\text{MHQC} = \log \hat{\sigma}_{p,T}^2 + [2 \log \log f(T)/f(T)]N,$$

where  $f(T) = \sum_{i=1}^T \kappa_{T-i} = T - T_e + 1$  is the effective sample size,  $T_e$  denotes the time index of the first observation of the effective sample,  $y(T_e), \dots, y(S), \dots, y(T)$ ,  $N$  is the number of functionally independent parameters, and  $\hat{\sigma}_{p,T}^2$  is the sample estimate of  $\sigma^2$ .

This chapter is organised as follows. Section 8.2 reviews subset AR modelling with a forgetting factor, and discusses AR order selection. Section 8.3 introduces two procedures – the bootstrap and the TRML estimation – to determine the value of the forgetting factor. Section 8.4 presents two illustrations to demonstrate the usefulness of the proposed procedures. A summary is provided in Section 8.5.

## 8.2 Subset AR Modelling with a Forgetting Factor

In recent years, an interesting development in time series analysis has been the theory of subset AR modelling. Empirical research has shown that it is impractical to ignore the possibility of zero coefficients in AR models, and the estimation and forecasting results could be very different if the presence of zero coefficients is allowed [see Penm and Terrell (1984a, 1984b), Penm *et al* (1997)]. This section presents the method which incorporates a forgetting factor into subset AR modelling.

To incorporate either a fixed or dynamic forgetting factor into subset AR modelling, the superscript  $T$  denotes the transpose, and (8.1) is re-written with the following representation:

$$\underline{u}_T + \underline{a}_p^T(I_s)U_{p,T-1} = \underline{\varepsilon}_T, \quad (8.3)$$

where  $I_s$  denotes an integer set with elements  $i_1, i_2, i_s, 1 \leq i_1 \leq i_2, \dots, \leq i_s$ , which represent zero coefficients in the subset AR model.

The variables in (8.3) for a given  $I_s$  can be defined as follows. Let

$$\underline{u}_T = [\sqrt{\kappa_0} y(T) \cdots \sqrt{\kappa_{T-p+1}} y(p+1) \cdots \sqrt{\kappa_{T-2}} y(2) \sqrt{\kappa_{T-1}} y(1)], \underline{a}_p^T = [a_1 \cdots a_p],$$

$$\underline{\varepsilon}_T = [\varepsilon(T) \cdots \varepsilon(1)] \text{ and}$$

$$U_{p,T-1} = \begin{bmatrix} \sqrt{\kappa_0} y(T-1) & \cdots & \sqrt{\kappa_{T-p-1}} y(p) & \cdots & \sqrt{\kappa_{T-2}} y(1) \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \sqrt{\kappa_0} y(T-p) & \cdots & \sqrt{\kappa_{T-p-1}} y(1) & \begin{array}{c} \nearrow \\ 0 \end{array} & \end{bmatrix},$$

then  $a_p^T(I_s)$  is formed by placing 0 in the  $(i_1, i_2, \dots, i_s)$ -th column elements of  $a_p^T$ .

In (8.3),  $a_p^T(I_s)$  can be estimated using the least squares method. However, there are  $p$  possible zero or non-zero coefficients in (8.3), which implies the existence of  $2^p$  possible subset AR models in the form of (8.3). To select the optimal specification of a subset AR model, Penm and Terrell (1984a) developed a tree search algorithm in conjunction with order selection criteria. This procedure can be utilised in a straightforward fashion to determine the optimal specification for (8.3).

As mentioned earlier, the forgetting factor has mainly been used in the analysis of non-stationary time-series based on AR model parameter estimation and order selection. The usefulness of order selection criteria in determining the specification of such a model was examined by Paulsen (1984) and Pötscher (1989). For a similar class of order selection criteria these researchers showed that results on consistency, which are valid in the stationary case, can be generalised to processes with roots which are on or within the unit circle. Empirical studies have been undertaken by Paulsen (1984), Goto *et al* (1995), and Penm *et al* (1997), to examine the efficiency of order selection criteria in determining the order for a system in the analysis of non-stationary time-series. The results indicate that the performance is broadly consistent with that of stationary systems reported by Penm

and Terrell (1984b). Given these results, the procedure developed by Penm and Terrell, together with order selection criteria, can be extended to identify the subset AR model for an integrated series. Thus the method presented above can be applied to the analysis of non-stationary series including a forgetting factor.

### **8.3 The Bootstrap and the TRML Estimation to Select the Forgetting Factor**

An important issue relating to the use of the forgetting factor in AR modelling is how to determine its value in various applications. The conventional methods are typically based on arbitrary or personal choices. Section 8.3.1 introduces the bootstrap to select the value of a fixed forgetting factor. Section 8.3.2 utilises the TRML, in conjunction with the bootstrap, to estimate the value of a dynamic forgetting factor.

#### **8.3.1 The Bootstrap for a Fixed Forgetting Factor**

The bootstrap is a statistical technique which permits the assessment of variability in an estimate using just the data at hand (Efron 1982). The idea is to approximate the theoretical distribution of a disturbance term by its empirical distribution and re-sample the original observations in a suitable way to construct ‘pseudo-data’ on which the estimator is based. Measures of variability, confidence intervals and even estimates of bias can then be obtained by repeating this process. The use of the bootstrap to evaluate forecasting models was first suggested by Peters and Freedman (1985).



To reduce computational complexity the bootstrap is applied to univariate AR modelling in this section. The approach can be extended to VAR, VECM and VRDL modelling.

As noted in Section 8.1 the observations  $y(t)$   $\{t=1, 2, \dots, T\}$  are available, this time-series is then fitted to Equation (8.1) with a fixed forgetting factor,  $\lambda$ . Forecasts can then be produced using the following equations:

$$\hat{y}(T+1) = -\underline{\theta}_{p,T}(I_s)[y(T) \cdots y(T-p+1)]^T,$$

$$\hat{y}(T+2) = -\underline{\theta}_{p,T}(I_s)[y(T+1) \cdots y(T-p+2)]^T,$$

$\vdots$

$$\hat{y}(T+i-1) = -\underline{\theta}_{p,T}(I_s)[\hat{y}(T+i-2) \cdots \hat{y}(T+i-1-p)]^T,$$

$$\hat{y}(T+i) = -\underline{\theta}_{p,T}(I_s)[\hat{y}(T+i-1) \cdots \hat{y}(T+i-p)]^T,$$

where  $\underline{\theta}_{p,T}(I_s)$  is the estimate of  $\underline{a}_p^*(I_s)$  using data up to  $y(T)$ , and  $\hat{y}(T+i)$ ,  $i=1, \dots, f$ , are the forecasts based on data up to  $y(T)$ .

Now the bootstrap can commence with the estimated  $\underline{\theta}_{p,T}(I_s)$ . For a large sample case the effect of  $\kappa_{T-t}$  declines as time  $t$  decreases from time  $T$ . At the point where  $t$  falls to  $T_e$  the effect of  $\kappa_{T-t}$  becomes negligible and diminishes to a value which is less than a tolerance  $\eta$ . Therefore a natural choice of the starting values to generate the pseudo-data would be the observations  $y(T_e - p)$ , ...,  $y(T_e - 1)$ . If the sample size is small, then  $T_e$  is irrelevant and should be equal to  $p+1$ . Thus the starting values would be the observations  $y(1), \dots, y(p)$ .

The estimated residuals can be obtained from the following equations:

$$\hat{\varepsilon}(t) = \sqrt{\kappa_{T-t}} y(t) - \sqrt{\kappa_{T-t}} \underline{\theta}_{p,T} (I_s) [y(t-1) \cdots y(t-p)]^T, t=T_e, \dots, T.$$

Consequently, the pseudo-data,  $y^*(T_e), \dots, y^*(T)$ , and the pseudo-future,  $y^*(T+1), \dots, y^*(T+f)$ , can be obtained from the following equations with re-sampled residuals. That is:

$$\begin{aligned} \sqrt{\kappa_{T-T_e}} y^*(T_e) &= -\sqrt{\kappa_{T-T_e}} \underline{\theta}_{p,T} (I_s) [y(T_e-1) \cdots y(T_e-p)]^T + \hat{\varepsilon}_{T_e}^*, \\ &\vdots \\ \sqrt{\kappa_1} y^*(T-1) &= -\sqrt{\kappa_1} \underline{\theta}_{p,T} (I_s) [y(T-2) \cdots y(T-p-1)]^T + \hat{\varepsilon}_{T-1}^*, \\ \sqrt{\kappa_0} y^*(T) &= -\sqrt{\kappa_0} \underline{\theta}_{p,T} (I_s) [y(T-1) \cdots y(T-p)]^T + \hat{\varepsilon}_T^*, \\ \sqrt{\kappa_{-1}} y^*(T+1) &= -\sqrt{\kappa_{-1}} \underline{\theta}_{p,T} (I_s) [y(T) \cdots y(T-p+1)]^T + \hat{\varepsilon}_{T+1}^*, \\ &\vdots \\ \sqrt{\kappa_{-f}} y^*(T+f) &= -\sqrt{\kappa_{-f}} \underline{\theta}_{p,T} (I_s) [y(T+f-1) \cdots y(T+f-p)]^T + \hat{\varepsilon}_{T+f}^*, \end{aligned}$$

where  $\sqrt{\kappa_{-i}}=1, i=0, \dots, f$ , and  $\hat{\varepsilon}_t^*, t=T_e, \dots, T+f$  is obtained by randomly drawing from  $\hat{\varepsilon}_s, s=T_e, \dots, T$ , with replacement. To reduce the effect of initial values, the first 100 outputs from a uniform-distributed random number generator are discarded.

The AR model can be re-estimated using the pseudo-data  $y^*(T_e), \dots, y^*(T)$ .

Likewise a set of pseudo-forecasts,  $\hat{y}^*(t), t=T+1, \dots, T+f$ , can be produced. The pseudo-errors,  $y^*(T+i) - \hat{y}^*(T+i), i=1, \dots, f$ , can also be calculated. Thus the distribution of the pseudo-errors can be used to approximate the distribution of the actual forecast errors.

The bootstrap procedure can be repeated many times. Now let  $\eta_i$  denote the standard error of the  $i$ -th forecast. Following Penm *et al* (1992) the overall standard error (OSE) is the average of the standard error of each forecast:

$$\text{OSE} = \sum_{i=1}^f (1/f) \eta_i .$$

The OSE is then used to assess the performance of the bootstrap procedure. The bootstrap procedure with a fixed forgetting factor is suitable for a model in which the underlying relationships between the variables involved change smoothly and gradually. Non-stationarity is not a concern, because the smaller is  $\lambda_{T-t}$ , the more rapid is the forgetting.

### 8.3.2 The TRML Estimation for a Dynamic Forgetting Factor

In this section the time-recursive maximum likelihood (TRML) estimation for updating of a subset AR model with a dynamic forgetting factor is formulated. The concept of a dynamic forgetting factor is introduced by Fortescue *et al* (1981) to avoid a 'blow up' of the covariance matrix of the estimates and subsequent unstable control.

Following Hannan and Deistler (1988), the time-update recursions for the TRML estimation with a fixed forgetting factor,  $\lambda(t)=\lambda$  can be described as follows.

Consider

$$\underline{\theta}_{p,t}^T = \underline{\theta}_{p,t-1}^T - \underline{g}_{p,t} e_{p,t},$$

where  $\underline{\theta}_{p,t}^T$  denotes the vector of coefficients estimated using data up to  $y(t)$ . Let

$$\underline{x}_{p,i} = [y(i) \ y(i-1) \cdots y(i-p+1)]^T. \text{ Then } e_{p,t} = y(t) + \sum_{i=1}^p \underline{\theta}_{p,t-1}^T \underline{x}_{p,t-1} \text{ is the}$$

prediction error,  $\underline{g}_{p,t} = P_{p,t-1} \underline{x}_{p,t} [\lambda + \underline{x}_{p,t}^T P_{p,t-1} \underline{x}_{p,t}]^{-1}$  is the Kalman gain vector, and

$$P_{p,t} = \frac{1}{\lambda} [P_{p,t-1} - \underline{g}_{p,t} \underline{x}_{p,t}^T P_{p,t-1}] \text{ is the inverse information matrix. The asymptotic}$$

memory length is defined as  $L=1/(1-\lambda)$ , which means that the information dies away with memory length,  $L$ .

In order to reduce computational complexity, Carayannis *et al* (1986) have presented a number of time-recursive algorithms based on the pre-windowed method for full-order AR modelling with a fixed forgetting factor. Toplis and Pasupathy (1988) have introduced a fast *a posteriori* error sequential technique (FAEST) type of algorithm for full-order AR modelling with a dynamic forgetting factor. These algorithms are theoretically equivalent to TRML estimation and are computationally efficient. However, they cannot be applied to subset AR modelling without modification. If zero coefficient values were assigned to the missing lags in subset AR models and these algorithms were applied in a straightforward fashion, the properties of subset AR models will disappear and the time-recursive algorithms would fail. For subset AR modelling with a fixed forgetting factor, Penm *et al* (1995, 1997) have developed an applicable algorithm for the time-update recursions. In the following, this algorithm is extended to

subset AR modelling with a dynamic forgetting factor. As the Kalman gain vector,  $\underline{g}_{p,t}$ , is efficiently updated at successive time instants, this algorithm can be regarded as a fast Kalman type of algorithm for subset AR modelling.

To introduce this time-recursive algorithm, a backward subset AR(p) model is considered as follows.

$$y(t-p) + \sum_{i=1}^p b_i(M_s)y(t-p+i) = \varepsilon(t, M_s), \{ b_i(M_s) = 0, \text{ as } i \in M_s, \} \quad (8.4)$$

where  $\varepsilon(t, M_s)$  is a zero mean Gaussian white noise disturbance term, and  $M_s$  represents an integer set with elements,  $m_1, m_2, \dots, m_s$ , denoting the zero coefficients,  $m_j = p - i_j$  and  $j = 1, 2, \dots, s$ . A reciprocal integer pair for a forward subset AR model and a backward subset AR model is a pair of (8.1) and (8.4). Figure 8.1 illustrates the reciprocal integer pairs for all subset AR models with a maximum order of 4, i.e.  $K=4$  and  $K \geq p$ . In this figure, numerals denote particular lags in a forward AR and underlined numerals represent such leads in a backward AR. Since the actual scheme of (8.4) may not be exactly of order  $K$ , the resulting estimates of  $b_p$  are denoted by  $b_{p,T}(k)$ , where  $T$  is the sample size.

**Figure 8.1**

**The reciprocal integer pairs of a subset AR model with K=4<sup>a</sup>**

---

Forward AR	Backward AR	Forward AR	Backward AR	Forward AR	Backward AR	Forward AR	Backward AR
1	<u>1</u>	12	<u>12</u>	123	<u>123</u>	1234	<u>1234</u>
2	<u>2</u>	13	<u>23</u>	124	<u>234</u>		
3	<u>3</u>	14	<u>34</u>	134	<u>134</u>		
4	<u>4</u>	23	<u>13</u>	234	<u>124</u>		
		24	<u>24</u>				
		34	<u>14</u>				

---

<sup>a</sup>Numerals represent particular lags in a forward AR and underlined numerals denote such leads in a backward AR.

Next, following Toplis and Pasupathy (1988), for observation  $i$  the following relations arise:

$$\underline{y}_{p,i}^T = [y(i) \ y(i-1) \cdots \ y(i-p)] \text{ where } y(j) = 0 \text{ for } j \leq 0 \quad \text{and}$$

$$\underline{z}_{p,T,i}^T = [\sqrt{\kappa_{T-i}(T)}y(i) \ \sqrt{\kappa_{T-i}(T-1)}y(i-1) \cdots \ \sqrt{\kappa_{T-i}(T-p)}y(i-p)],$$

$$\text{with } \kappa_a(T) = \begin{cases} \prod_{i=1}^a \lambda(T-i) & a \geq 1 \\ 1 & a < 1 \end{cases},$$

$$\Gamma_{p,T} = \begin{bmatrix} \sqrt{\lambda(T)/\lambda(T-1)} & & & & \\ & \sqrt{\lambda(T)/\lambda(T-2)} & & & 0 \\ & & \ddots & & \\ 0 & & & \ddots & \\ & & & & \sqrt{\lambda(T)/\lambda(T-p)} \end{bmatrix},$$

and

$$\Omega_{p,T} = \begin{bmatrix} \sqrt{\lambda(T-p)/\lambda(T)} & & & & \\ & \sqrt{\lambda(T-p)/\lambda(T-1)} & & & 0 \\ & & \ddots & & \\ 0 & & & \ddots & \\ & & & & \sqrt{\lambda(T-p)/\lambda(T-p+1)} \end{bmatrix}.$$

Thus, for a forward AR(p,  $I_s$ ) fitted with this data set, the following relations emerge:

$$R_{p,T}(I_s) = \sum_{i=1}^T \underline{z}_{p,T,i}(L_s) \underline{z}_{p,T,i}^T(L_s),$$

$$R_{p,T}(I_s) \begin{bmatrix} 1 \\ \underline{a}_{p,T}^T(I_s) \end{bmatrix} = \begin{bmatrix} v_{p,T}(I_s) \\ 0 \end{bmatrix}$$

and the forward prediction error  $e_{p,T+1}(I_s) = [1 \quad \underline{a}_{p,T}^T(I_s) \Gamma_{p,T}(I_s)] \underline{y}_{p,T+1}(L_s)$ , where

$\underline{a}_{p,T}^T = [a_{p,T}(1) \quad a_{p,T}(2) \quad \cdots \quad a_{p,T}(p)]$  with  $a_{p,T}(k)$  denotes the estimate of  $a_k$  in

(A.4.1), and  $v_{p,T}$  is the forward residual sum of squares. If  $L_s$  represent an integer set with elements  $l_j, j=1, \dots, s$ , and  $l_j=i_j+1$ , then  $\underline{z}_{p,T,i}(L_s)$  and  $\underline{y}_{p,i}(L_s)$  are formed by removing the  $(l_1, \dots, l_s)$ -th row elements of  $\underline{z}_{p,T,i}$  and  $\underline{y}_{p,i}$  respectively,  $\underline{a}_{p,T}^T(I_s)$  is formed by removing the  $(i_1, i_2, \dots, i_s)$ -th column elements of  $\underline{a}_{p,T}^T$ , and  $\underline{\Gamma}_{p,T}(I_s)$  is formed by removing the  $(i_1, i_2, \dots, i_s)$ -th rows and columns of  $\underline{\Gamma}_{p,T}$ .

In addition, for the corresponding backward AR(p,  $M_s$ ), the following relations arise:

$$\underline{b}_{p,T}^T = [b_{p,T}(p) \cdots b_{p,T}(1)], \quad R_{p,T}(M_s) \begin{bmatrix} \underline{b}_{p,T}^T(M_s) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{v}_{p,T}(M_s) \end{bmatrix},$$

$$D_{p,T}(M_s) = \sum_{i=1}^T \underline{z}_{p-1,T,i}(M_s) \underline{z}_{p-1,T,i}^T(M_s),$$

$$\underline{g}_{p,T}(M_s) = D_{p,T}^{-1}(M_s) \underline{y}_{p-1,T}(M_s),$$

$$\beta_{p,T}(M_s) = 1 - \underline{y}_{p-1,T}^T(M_s) \underline{g}_{p,T}(M_s),$$

and the backward prediction error  $\bar{e}_{p,T+1}(M_s) = [\underline{b}_{p,T}^T(M_s) \Omega_{p,T}(M_s) \quad 1] \underline{y}_{p,T+1}(M_s)$ ,

where  $\bar{v}_{p,T}(I_s)$  is the backward residual sum of squares,  $\underline{b}_{p,T}^T(M_s)$  is formed by removing the  $(p+1-m_1), \dots, (p+1-m_s)$ 'th column element of  $\underline{b}_{p,T}^T$ , and  $\underline{y}_{p-1,i}(M_s)$  is formed by removing the  $(p+1-m_1), \dots, (p+1-m_s)$ 'th row element of  $\underline{y}_{p-1,i}$ . Thus



$\underline{\Omega}_{p,T}(\mathbf{M}_s)$  is formed by removing the  $(p+1-m_1), \dots, (p+1-m_s)$ 'th rows and columns of

$\underline{\Omega}_{p,T}$ . Of note,  $p+1-m_j=1+i_j=l_j$ , and thus

$$\underline{y}_{p-1,i}(\mathbf{M}_s) = \underline{y}_{p-1,i}(\mathbf{L}_s),$$

and

$$\mathbf{R}_{p,T}(\mathbf{M}_s) = \mathbf{R}_{p,T}(\mathbf{I}_s).$$

Next, the new information  $y(T+1)$  is assumed to be available and a backward  $\text{AR}(p+1, N_s)$  model fitted with the sample,  $y(1), \dots, y(T+1)$  is considered, where  $n_j$  of  $N_s$  has the relation  $n_j=m_j+1$ . As a result:

$$\mathbf{D}_{p+1,T+1}(\mathbf{N}_s) = \sum_{i=1}^T \underline{z}_{p,T+1,i}(\mathbf{M}_s) \underline{z}_{p,T+1,i}^T(\mathbf{M}_s),$$

$$\underline{g}_{p+1,T+1}(\mathbf{N}_s) = \mathbf{D}_{p+1,T+1}^{-1}(\mathbf{N}_s) \underline{y}_{p-1,T}(\mathbf{M}_s),$$

$$\beta_{p+1,T+1}(\mathbf{N}_s) = 1 - \underline{y}_{p,T+1}^T(\mathbf{M}_s) \underline{g}_{p+1,T+1}(\mathbf{M}_s).$$

When a backward  $\text{AR}(p, N_s)$  model fitted with the sample,  $y(1), \dots, y(T)$  is considered, the following relations emerge:

$$D_{p+1,T+1} (N_s) = \sum_{i=1}^T \underline{z}_{p,T+1,i} (M_s) \underline{z}_{p,T+1,i}^T (M_s),$$

$$\underline{g}_{p+1,T+1} (N_s) = D_{p+1,T+1}^{-1} (N_s) \underline{y}_{p-1,T} (M_s),$$

$$\beta_{p+1,T+1} (N_s) = 1 - \underline{y}_{p,T+1}^T (M_s) \underline{g}_{p+1,T+1} (M_s).$$

Suppose  $\underline{a}_{p,T} (I_s)$ ,  $\underline{y}_{p,T+1} (I_s)$ ,  $\beta_{p,T} (N_s)$ ,  $\underline{g}_{p,T} (N_s)$ ,  $\underline{b}_{p,T} (M_s)$ ,  $v_{p,T} (I_s)$ ,  $\bar{v}_{p,T} (M_s)$  and  $\lambda(t)$  are available, then the proposed time-update recursions are as follows:

Forward time-update recursions from T to T+1

$$e_{p,T+1} (I_s) = [1 \ \underline{a}_{p,T}^T (I_s) \Gamma_{p,T} (I_s)] \underline{y}_{p,T+1} (I_s) \quad (8.5.1)$$

$$\epsilon_{p,T+1} (I_s) = e_{p,T+1} (I_s) \beta_{p,T} (N_s) \quad (8.5.2)$$

$$\underline{a}_{p,T+1} (I_s) = \underline{a}_{p,T} (I_s) \Gamma_{p,T} (I_s) - \underline{g}_{p,T} (N_s) e_{p,T+1} (I_s) \quad (8.5.3)$$

$$v_{p,T+1} (I_s) = \lambda(T) v_{p,T} (I_s) + \epsilon_{p,T+1} (I_s) e_{p,T+1} (I_s) \quad (8.5.4)$$

$$\beta_{p+1,T+1} (N_s) = \beta_{p,T} (N_s) - v_{p,T+1}^{-1} (I_s) \epsilon_{p,T+1}^2 (I_s) \quad (8.5.5)$$

$$\underline{g}_{p+1,T+1} (N_s) = \begin{bmatrix} 0 \\ \underline{g}_{p,T} (N_s) \end{bmatrix} + \begin{bmatrix} 1 \\ \underline{a}_{p,T+1} (I_s) \end{bmatrix} v_{p,T+1}^{-1} (I_s) \epsilon_{p,T+1} (I_s) \quad (8.5.6)$$

$$\text{partition } \underline{g}_{p+1,T+1}(N_s) = \left[ \frac{d}{d} \right]$$

$$\bar{e}_{p,T+1}(M_s) = \lambda(T-p)d\bar{v}_{p,T}(M_s)\beta_{p+1,T+1}^{-1}(N_s) \quad (8.5.7)$$

$$\beta_{p,T+1}(M_s) = \beta_{p+1,T+1}(N_s) - \bar{e}_{p,T+1}(M_s)d \quad (8.5.8)$$

$$\bar{e}_{p,T+1}(M_s) = \bar{e}_{p,T+1}(M_s)\beta_{p,T+1}(M_s) \quad (8.5.9)$$

$$\underline{g}_{p,T+1}(M_s) = \frac{d - \underline{b}_{p,T}(M_s)\Omega_{p,T}(M_s)d}{1 - \bar{e}_{p,T+1}(M_s)d} \quad (8.5.10)$$

$$\bar{v}_{p,T+1}(M_s) = \lambda(T-p)\bar{v}_{p,T}(M_s) + \bar{e}_{p,T+1}(M_s)\bar{e}_{p,T+1}(M_s) \quad (8.5.11)$$

$$\underline{b}_{p,T+1}(M_s) = \underline{b}_{p,T}(M_s)\Omega_{p,T}(M_s) - \underline{g}_{p,T+1}(M_s)\bar{e}_{p,T+1}(M_s) \quad (8.5.12)$$

If there is a consecutive set of  $k$  deleted leads beginning at lead 1 in the backward AR( $p, M_s$ ), then the following relations arise:

$$\underline{g}_{p,T+1}(M_s) = \underline{g}_{p-k,T+1}(M_k) \text{ and } \beta_{p,T+1}(M_s) = \beta_{p-k,T+1}(M_k), \quad (8.5.13)$$

where  $M_s$  contains  $m_1, \dots, m_k, \dots, m_s, m_j=j, j=1, 2, \dots, k$ , and  $M_k$  contains  $m_{k+1-k}, \dots, m_s-k$ .

If there is a consecutive set of  $k$  deleted leads beginning at lead  $p$  in the backward  $AR(p, N_s)$ , then

$$\underline{g}_{p,T}(N_s) = \underline{g}_{p-k,T-k}(N_k) \text{ and } \beta_{p,T}(N_s) = \beta_{p-k,T-k}(N_k), \quad (8.5.14)$$

where  $N_s$  contains  $n_1, \dots, n_{s-k}, n_{s-k+1}, \dots, n_s, n_{s+1-j}=p+1-j, j=1, 2, \dots, k$ , and  $N_k$  contains  $n_1, \dots, n_{s-k}$ .

For full-order AR models with the fixed forgetting factor, that is all  $I_s, M_s, N_s$  are empty sets and  $\lambda(t) = \lambda$ , the proposed recursions become the most efficient version of the fast Kalman algorithm presented in Carayannis *et al* (1986).

Following Cho *et al* (1991), the dynamic forgetting factor,  $\lambda(t)$ , is selected to satisfy the following condition:

$$\lambda(t) = 1 - \frac{1}{L(t)}, \text{ where } L(t) = \frac{\sigma_e^2}{q_t(1 - \lambda_{\max})}, \quad (8.6)$$

where  $\sigma_e^2$  is the expected measurement noise variance based on real knowledge of the process,  $\lambda_{\max}$  is the upper limit value of the dynamic forgetting variable. The extended prediction error variance which accounts for the non-stationarity of the signal is defined as

$$q_t = \frac{1}{M} \sum_{i=0}^{M-1} e_{p,t-i}^2, \quad (8.7)$$

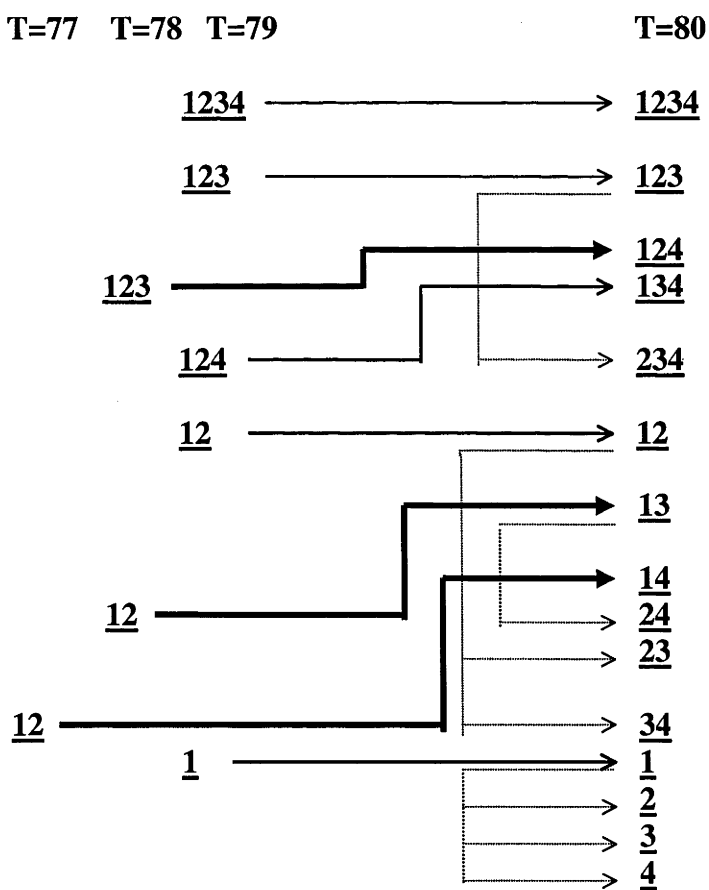
where  $M$  is chosen to minimise the effect of a spurious large additive prediction error. Once the value of  $\lambda(t)$  is determined, this value and its associated forgetting profile are incorporated into the proposed time-update recursions. As described in Section 8.1, the optimum subset AR model is then identified using MHQC.

Suppose the above time-update algorithm is undertaken for a subset AR model with  $K=4$ , and  $\underline{a}_{p,T}(\mathbf{I}_s)$ ,  $\underline{y}_{p,T+1}(\mathbf{I}_s)$ ,  $\beta_{p,T}(\mathbf{N}_s)$ ,  $\underline{g}_{p,T}(\mathbf{N}_s)$ ,  $\underline{b}_{p,T}(\mathbf{M}_s)$ ,  $v_{p,T}(\mathbf{I}_s)$ ,  $\bar{v}_{p,T}(\mathbf{M}_s)$  and  $\lambda(t)$  are all available. One possible initialisation of the recursions is to carry out the direct method for an initial block of the sample set. By employing the equations (8.5.1)–(8.5.12),  $\underline{a}_{p,T+1}(\mathbf{I}_s)$ ,  $\underline{b}_{p,T+1}(\mathbf{M}_s)$ ,  $v_{p,T+1}(\mathbf{I}_s)$  and  $\bar{v}_{p,T+1}(\mathbf{M}_s)$  can be obtained for each reciprocal integer pair of the AR( $p, \mathbf{I}_s$ ) and AR( $p, \mathbf{M}_s$ ) at  $T+1$ . Equation (8.5.13) indicates that both  $\underline{g}_{p,T+1}(\mathbf{N}_s)$  and  $\beta_{p,T+1}(\mathbf{N}_s)$  for certain backward autoregressive models also arise from another backward autoregression. In Figure 8.2, dashed lines provide the illustration. Thus,  $\underline{g}_{p,T+1}(\mathbf{N}_s)$  and  $\beta_{p,T+1}(\mathbf{N}_s)$ , corresponding to the backward autoregressions including lead 1, are required to carry out the recursions. However from (8.5.14) both  $\underline{g}_{p,T}(\mathbf{N}_s)$  and  $\beta_{p,T}(\mathbf{N}_s)$  for certain backward autoregressive models also arise from another backward autoregression. Solid thick lines in Figure 8.2 illustrate the recursions using (8.5.14) and solid thin lines illustrate the recursions without using (8.5.14). Figure 8.2 also indicates how to acquire the Kalman gain vector of each reciprocal integer pair of the forward and the backward autoregressions with  $K=4$  from the time-update recursions. Indeed, after the initialisation is completed,  $\underline{g}_{p,T}(\mathbf{N}_s)$  for

the next time-recursion can be obtained. Thus the algorithm can be carried out continuously.

### Figure 8.2

**Recursions to acquire  $\underline{g}$  and  $\beta$  for each reciprocal integer pair of the forward and backward AR models (K=4)<sup>a</sup>**



<sup>a</sup> Only the backward AR of each pair are listed. Solid thick lines represent the  $[\underline{g}_{p,T}(N_s), \beta_{p,T}(N_s)]$  to  $[\underline{g}_{p,T+1}(M_s), \beta_{p,T+1}(M_s)]$  recursions using (8.5.14), Solid thin lines represent the  $[\underline{g}_{p,T}(N_s), \beta_{p,T}(N_s)]$  to  $[\underline{g}_{p,T+1}(M_s), \beta_{p,T+1}(M_s)]$  recursions not using (8.5.14), and dashed lines denote  $g_{p,T}(M_s) = g_{p-k,T}(M_k)$  and  $\beta_{p,T}(M_s) = \beta_{p-k,T}(M_k)$ .

## 8.4 Case Studies

This section presents two case studies to demonstrate the effect of the forgetting factor method in subset AR modelling on *ex ante* forecasting of non-stationary time-series. Section 8.4.1 uses the bootstrap to examine the fixed forgetting factor on three real exchange rates. Section 8.4.2 uses the TRML, in conjunction with the bootstrap, to test the dynamic forgetting factor.

### 8.4.1 Use of the Bootstrap to Determine the Value of a Fixed Forgetting Factor

In this case study the bootstrap is utilised to determine the value of a fixed forgetting factor with a given sample. The effect of the fixed forgetting factor on forecasting is examined. The exchange rates under study are the bilateral exchange rates between the US Dollar and three other currencies comprising the German Deutschmark, the UK Sterling and the Korean Won. A real exchange rate series provides a measure of the international competitiveness of the economy, and a stable real exchange rate series means that business firms face minimal economic risk. Therefore real exchange rates are studied.

The real exchange rates are obtained by adjusting the monthly average nominal bilateral exchange rates by the inflation rate differentials approximated by movements in the wholesale price indexes in the respective countries. The sample period covers January 1975 to December 1993, a total of 228 observations. All the series are in logarithms and are obtained from DataStream™. Figure 8.3

presents the real exchange rates for the USA with Germany, United Kingdom and Korea. The non-stationarity in the real exchange rate series was first examined, using the augmented Dickey-Fuller (ADF) unit root test. The results indicate non-stationarity in each of the real exchange rate series.

To decide how best to explain the movements of these series, which may be slowly evolving, the forecasting performance is compared using subset AR modelling, both with the forgetting factor and without the forgetting factor. If the series are better described by subset AR models using the forgetting factor, theorems based on stationarity would have less meaning. The findings reported below indicate consistently improved forecasting performance from the AR models using the forgetting factor. To briefly illustrate the potential use of the forgetting factor for improved forecasting performance in time-series which are non-stationary, only the bootstrap procedure is used.

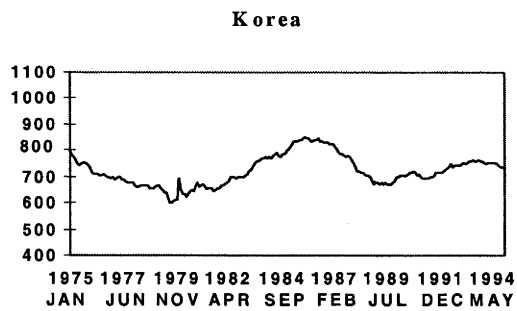
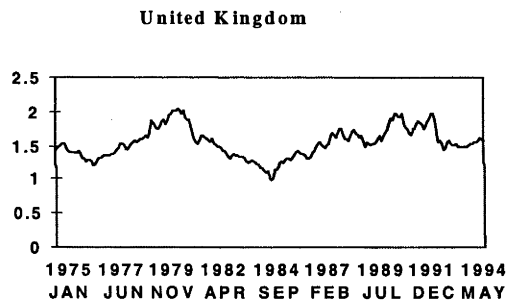
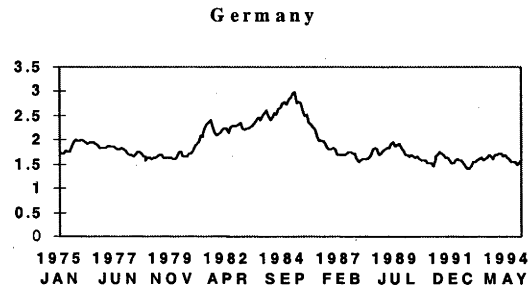
To determine the specifications of the subset AR models, the Penm and Terrell (1984a) procedure is utilised with the forgetting factor incorporated in each of the AR models. As part of the approach, a forgetting process is introduced. In this process, 'forgetting' occurs from observation 148 to observation 1 (that is,  $T-S=80$ , where  $T=228$  and  $S=148$ ).



**Figure 8.3**

**Real exchange rates for the USA with Germany, United Kingdom and  
Korea, monthly: January 1975 to December 1993**

---



Another forgetting process is also carried out. In this process, the last 30 observations are kept unchanged and the effect of  $\lambda_{T-t}$  has been assumed to be negligible when  $\lambda_{T-t}$  is less than the tolerance  $10^{-7}$ . The subset AR specifications with different forgetting factors are determined. Subsequently six *ex ante* forecasts are conducted. The subset AR specifications determined, and their associated percentage of improvement, are consistent with the results obtained in the T-S=80 case. The choices of 30 and 80 are for illustrative purposes only. A number of different values are assigned to the forgetting factor. The determined specifications for each AR model with different forgetting factors are presented in Table 8.1 These AR specifications are determined using MHQC.

It is interesting that in most cases the subset AR specifications determined with the forgetting factor are different from those without it. As the value of the forgetting factor is gradually reduced, changes to the specifications are also evident.

**Table 8.1**  
**Order of the subset AR models determined using MHQC**

Forgetting Factor	German Deutschmark Order	UK Pound Order	Korean Won Order
1.0	(1)	(1 2 3)	(3 7 9 10)
0.99	(1)	(1 2 3)	(3 7 9 10)
0.97	(1)	(1 2 3)	(1 2 3 8 9 11)
0.95	(1 2)	(1 2 5)	(1 2 3 8 9 11)
0.93	(1 5)	(1 2 5)	(1 2 3 8 9 11)
0.90	(1 2)	(1 2 5 10)	(1 2 3 8 9 11)

To examine the effects on forecasting the root mean squared error (RMSE) is computed, for six one-period-ahead *ex ante* forecasts outside the observed sample data respectively, generated by these subset AR models. Using the AR model without the forgetting factor as the baseline, the percentage of improvement (or deterioration) is calculated for each subset AR model with the forgetting factor. The percentages are presented in Table 8.2 under the column ‘Data’.

**Table 8.2**  
**Improvements<sup>a</sup> over *ex ante* forecasts using the forgetting factor**

	German Deutschmark		UK Pound		Korean Won	
Forgetting Factor	Data	Bootstrap	Data	Bootstrap	Data	Bootstrap
	%	%	%	%	%	%
0.99	0.0	4.6	-1.2	-1.5	0.3	0.0
0.97	12.6*	13.7*	-1.6	1.7	15.7*	40.0*
0.95	6.7	11.7	7.4	14.2	14.2	29.7
0.93	6.7	7.7	8.7*	35.8*	12.1	23.8
0.90	7.3	7.0	4.5	30.1	10.3	23.2

<sup>a</sup>The improvements are expressed as percentages of the results when a forgetting factor is not included. The bootstrap results are based on 100 replications. The symbol \* denotes the maximum improvement.

As portrayed in Table 8.2, the improvement (or deterioration) differs across cases. This creates a need to apply the bootstrap to determine the value of the forgetting factor. In Table 8.2 the bootstrap results based on 100 replications are also presented (using the label ‘Bootstrap’). Examination of the pseudo-errors across

each replication in this exercise shows very stable results, indicating that 100 replications are sufficient. As the OSE is employed to approximate the real forecast errors, the forgetting factor which produces the smallest OSE is expected to bring about the most significant improvement. For each case the improvement is also presented as a percentage of the OSE obtained from the AR model without the forgetting factor. Interestingly the bootstrap results are very consistent with those obtained using real data. That is, in almost all cases, the forgetting factor which generates the smallest OSE also produces the smallest RMSE in forecasting using real data. Also the bootstrap ensures more significant improvements in the forecasting performance. The outcome generally indicates that the bootstrap is a reliable procedure for determining the value of the forgetting factor in subset AR modelling.

However as the number of replications for the bootstrap increases (beyond 100), the execution time required becomes very considerable, and the associated costs are unaffordable. Clearly, a trade-off is required. Although the algorithm of the bootstrap is superior in the improvement of forecasting performance, some sacrifice in computational efficiency is unavoidable.

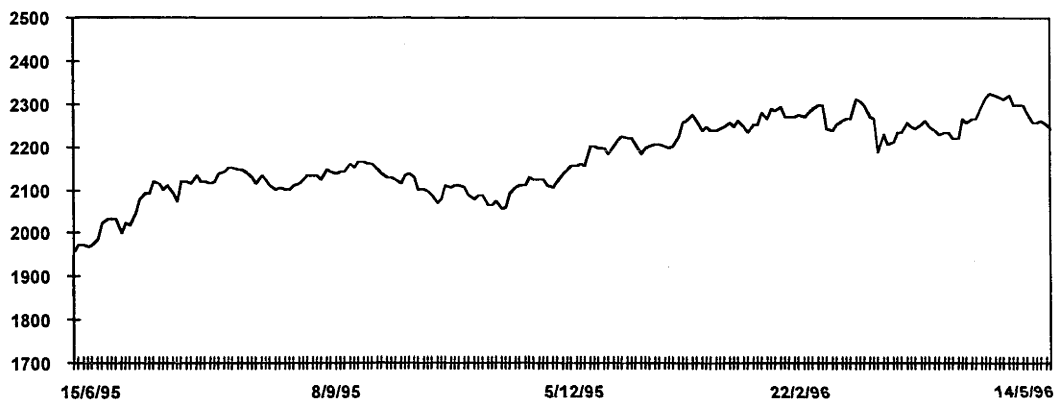
#### **8.4.2 Use of the TRML to Estimate the Value of a Dynamic Forgetting Factor**

In this second study, the daily All Ordinaries Index (AOI) sample covering the period 15 June 1995 to 14 May 1996, a total of 211 observations. The index series is in logarithms and presented in Figure 8.4. Since the series exhibits varying periodic behaviour and is non-stationary, it is preferable to be assessed by subset

time series models, which will be selected sequentially [see Penm *et al* (1995)]. A linear trend exists in this series and is removed before the forgetting factor analysis is conducted.

**Figure 8.4**

**The All Ordinaries Index levels series over the period 15 June 1995 to 14 May 1996**



To conduct the TRML estimation, an initial subset AR model with a forgetting factor is required. To determine this initial model, the bootstrap procedure described in Section 8.3.1 seems to be a reasonable choice. Applying the procedure over the first 195 observations, the results indicate that the specification of the initial subset AR model is (1, 5, 7), and the value of the forgetting factor is 0.99. The value of  $\hat{\sigma}_{p,T}^2$  is also calculated from this estimation.

Given these results, the TRML estimation for  $T+1=196$  is then conducted. The selection of the dynamic forgetting factor depends on the value of  $\lambda_{\max}$ ,  $\lambda_{\min}$ ,  $M$  and  $\hat{\sigma}_e^2$ . As the series under investigation is non-stationary after trend removal, an upper bound at  $\lambda_{\max} = 0.9995$  is chosen. To prevent  $\lambda(t)$  from becoming negative,  $\lambda_{\min}$  is set at 0.75. As reported in Cho *et al* (1991), the value of  $M$  is small enough not to obscure the non-stationarity of the signal. However if this value is too small the effect of a spurious large additive prediction error would become significant. This adverse effect leads to a wild fluctuation of the dynamic forgetting factor to be used in the next recursion of TRML estimation. These results indicate that  $M$  should be larger than 5 both to achieve a smooth updating of the dynamic forgetting factor and to prevent a large spurious noise error from creating the calculation of a misleading dynamic forgetting factor. Also the larger that  $M$  is, the higher is the likelihood of over-averaging conducted by  $q_t$ . Subsequently the non-stationarity of the signal is obscured. Therefore the value of  $M$  is set at 6. For a stationary time series, Cho *et al* (1991) indicate that  $\sigma_e^2$  approaches  $q_t$ . In this illustration the quantity  $\hat{\sigma}_e^2$  is set at  $10^{-4}$ , which is approximated by averaging the squared residuals of the initial model [see Toplis and Pasupathy (1988)].<sup>34</sup>

The value of  $\lambda(196)$  is calculated at 0.975 (Table 8.3). This value is marginally different from that determined in the initial model. This dynamic forgetting factor is then incorporated into the proposed TRML to select the updated subset AR model. The order specification of this updated model remains as (1, 5, 7). To

---

<sup>34</sup> Fortunately, the performance of the proposed algorithm appears to be insensitive to this quantity. Too small a value of  $\hat{\sigma}_e^2$  may lead to a corresponding growing residual, which is not obtained.

examine the effects on forecasting, the RMSE for six forward forecasts is again calculated. Compared with the subset AR model not utilising a forgetting factor, Table 8.3 shows that a 36.4 percent improvement in forecasting performance is achieved by using this updated model utilizing the forgetting factor.

This time-update is also undertaken for observations  $y(t)$ ,  $t=197, \dots, 205$ . The values of the dynamic forgetting factor,  $\lambda(t)$ ,  $t=197, \dots, 205$ , the updated subset AR models, and the improvements in forecasting performance, are presented in Table 8.3.

For the purpose of comparison, the bootstrap is also conducted in a time-update fashion for the same sample period. As previously presented, a range of possible values are first assigned to the fixed forgetting factor, and each of the associated subset AR models is determined using the algorithm proposed by Penm and Terrell (1984a). In the course of the bootstrap, 100 replications are performed for each case. In addition, 'forgetting' is set to begin immediately, that is  $S=T$ . To assess forecasting performance the RMSE for six forward forecasts in each case is also computed. These results are also presented in Table 8.3.

**Table 8.3****The TRML estimation and the bootstrap results for the AOI**

The TRML Estimation				The Bootstrap Estimation		
t	Order	Forgetting factor	Improvement <sup>a,b</sup>	Order	Forgetting factor	Improvement <sup>a</sup>
196	(1 5 7)	0.975	36.4%	(1 5 7)	0.99	46.8%
197	(1 5 7)	0.988	46.3%	(1 5 7)	0.99	45.2%
198	(1 4 7)	0.999	63.8%	(1 4 7)	0.99	63.3%
199	(1 4 7)	0.999	63.4%	(1 4 7)	0.99	63.2%
200	(1 5 7)	0.999	64.9%	(1 5 7)	0.99	65.8%
201	(1 5 7)	0.994	68.3%	(1 5 7)	0.97	70.0%
202	(1 5 7)	0.996	58.7%	(1 5 7)	0.97	61.1%
203	(1 5 7)	0.999	39.1%	(1 5 7)	0.98	44.5%
204	(1 5 7)	0.999	32.4%	(1 5 7)	0.98	38.0%
205	(1 5 7)	0.985	23.4%	(1 5 7)	0.97	30.3%

<sup>a</sup> Expressed as percentages of the results when a forgetting factor is not included.

<sup>b</sup>  $\lambda = .99$  for the initial period.

Interestingly, the subset AR models determined in this fixed forgetting factor analysis exhibit the same specifications as those selected in the TRML procedure. This indicates that the value of the forgetting factor is the main contributor to forecasting improvement. It is also noteworthy that the forgetting profiles are different in these two procedures, even though the forgetting factors determined in these two procedures have the same value. It is also observed that  $\lambda$  is generally higher than that for exchange rates.

In theory the inclusion of a dynamic forgetting factor in AR modelling, using the TRML estimation, should have the potential to outperform the bootstrap in a time-update fashion. This is because the former possesses a higher degree of flexibility in terms of the 'forgetting' process, which should lead to superior modelling, and



hence improved forecasts. However this is not fully supported by the results (see Table 8.3), although noting the small sample size. One possible explanation is that the TRML estimation relies on asymptotic theory, while the bootstrap is based on an empirical distribution associated with the data sample for each series.

A comparison of the computer-execution time between the TRML estimation and the bootstrap in a time-update fashion is then conducted. The execution time required for the specification of subset AR models, and the associated values of the forgetting factor determined by each proposed procedure, are recorded. The results of the comparison indicate that on average the execution time for the bootstrap is around five times longer than that for the TRML estimation. As the number of replications for the bootstrap increases and is larger than the current practice of 100 replications, the execution time required becomes substantial and the associated costs are unaffordable. Although the algorithm of the bootstrap is superior in conceptual simplicity and easier to maintain, sacrifice in computational efficiency is unavoidable.

## **8.5 Summary**

In this chapter, two procedures were presented which can be used to determine the value of the forgetting factor in subset AR modelling. The first was based on the bootstrap to select the value of a fixed forgetting factor. The second used the TRML estimation, in conjunction with the bootstrap, to estimate the value of a dynamic forgetting factor. Two case studies were conducted. First, the bootstrap procedure, applied to real exchange rates, demonstrated that it is reliable in

determining the value of the fixed forgetting factor. Second, the TRML estimation procedure, applied to the AOI, showed that it is effective in deciding the value of the dynamic forgetting factor. The results consistently indicated that the forgetting factor, so determined, can bring about significant improvements in forecasting.

## CHAPTER 9

### SUMMARY AND CONCLUSIONS

In this thesis advanced non-linear statistical and econometric models have been combined and developed, in the context of international financial markets. The goal of this thesis was to derive new modelling techniques, and improve our understanding of the price behaviour of financial markets, particularly foreign exchange markets, thereby adding to the stock of knowledge on these topics.

VAR models are a useful tool to analyse relations within and between financial markets. In such models, there are structures that require zero entries in the coefficient matrices. The main contribution of the thesis has been the development of parsimonious ZNZ patterned vector time-series models, which allow for possible zero entries in coefficient matrices. These patterned models have been applied to a wide range of problems arising in exchange rate applications.

More specifically, this thesis has shown that a direct extension of the use of the Yule-Walker relations for fitting ZNZ patterned VAR models is inconsistent with statistical procedures as the resultant estimated variance-covariance matrix of the white noise disturbance process becomes non-symmetric. This inconsistency can hinder efforts to effectively test financial theories. Chapter 3 provided a consistent adjustment which fits with statistical theory and allows necessary testing. The adjustment is consistent with statistical procedure in theory and has the advantage of computational efficiency and reliability.

Chapter 4 then demonstrated that ZNZ patterned VAR models could be used as a basis for detecting Granger causality, Granger non-causality and indirect causality for stationary vector financial time-series. The identified Granger causal relations derived from the ZNZ patterned VAR models with unit roots were shown to be identical with the causal relations identified from those derived from the equivalent ZNZ patterned VECM.

Both Chapters 3 and 4 focused on an effective and efficient algorithm to select the optimal ZNZ patterned cointegrating and loading vectors in a ZNZ VECM framework for an  $I(1)$  system. The algorithm can be applied to a higher order integrated systems. The proposed algorithm is simple to use and leads to an efficient analysis of the cointegrating relationships in vector financial time-series. To demonstrate the usefulness of this algorithm, applications to financial markets were presented.

In Chapter 5 the procedure for selecting the optimal ZNZ VAR was utilised to investigate direct Granger causal relations between the money supply and the Euro exchange rate. These two variables have important linkages and their relationship has not previously been examined in the context of the European Monetary Union. The empirical results are consistent with both theory and prior evidence.

In Chapter 6 the relevance of PPP was tested using various exchange rate series. While empirical testing of the PPP hypothesis has received significant attention in the literature, the introduction of the cointegration theory under the framework of ZNZ patterned VECM modelling provides new opportunities to accommodate both long-term and dynamic responses. Three unit root tests were utilised for fourteen real bilateral exchange rates under examination. PPP was consistently accepted or rejected for only four exchange rates. For the remaining ten, PPP was accepted using the Kwiatkowski *et al* (1992) procedure, but rejected by the augmented Dickey-Fuller test and the Phillips-Perron test. These conflicting results are attributed mainly to the sensitivity of statistical procedures used in testing PPP to the associated null hypothesis. In an alternative approach, the PPP relationship was estimated in a ZNZ VECM framework. The determined specification gives support for the presence of PPP for seven out of fourteen exchange rates, and indicates that reliance on such a long-term relationship is influenced by the current assessment of the short-term movements in the exchange rate. The outcome suggests that emphasis on the specification could be useful in improving the test for PPP.

Further, two case studies using an I(2) algorithm were analysed. The first case study examined PPP in the Australian market. The findings confirm support for the necessary condition of the PPP hypothesis for the bilateral exchange rate between the Australian and US Dollars. They also indicate that direct or indirect causality exists among the nominal exchange rate, domestic and foreign price levels. The second case study dealt with the inter-relationships between the stock market, money supply and inflation. These results confirm that money supply is an

independent source of financial and economic disturbance, and money supply impacts on stock prices through inflationary pressures. These findings are generally consistent with both economic theory and prior evidence.

The use of computer-intensive statistical methods including neural networks, forgetting factor, and bootstrapping in financial time-series research has offered opportunities to provide new insights into financial time-series. Such methods are flexible in accounting for potentially complex non-linear relationships not fully captured by linear regression methods. In Chapter 7 a numerically robust lattice-ladder learning algorithm was developed. At each point of time it reassess the order and parameter estimation for a subset time-series system using ZNZ patterned polynomial neural networks. The proposed lattice-ladder learning algorithm for extended neural networks avoids cumbersome matrix inversion and results in better numerical accuracy and importantly captures slow evolution in the parametric model structure.

Chapter 8 introduced and developed the forgetting factor method in subset AR modelling. In contrast to the conventional methods which are mostly based on arbitrary or personal choices, this chapter proposed two procedures to determine the value of the forgetting factor. The first method used the bootstrap to select the value of a fixed forgetting factor. The second method utilised both the bootstrap and the TRML estimation to determine the value of a dynamic factor. The findings show that the forgetting factor so determined can improve the forecasting performance.

The major theoretical contributions and empirical findings of this thesis have been presented in Chapters 1 to 8. The thesis also contains a number of appendices. The purpose of appendices is to provide details that support the models, contributions and findings included in the chapters. Appendix A supplements Chapter 3, Appendix B supplements both Chapters 4 and 6, and Appendices C and D supplement Chapters 7 and 8 respectively.

Some of the approaches developed in the thesis, such as parsimonious patterned modelling and computer-intensive statistical methods, are novel in nature. The combination of these advanced approaches have provided forecasting improvements in the performance of vector time-series models. The main theoretical contributions of the thesis are ZNZ patterned VAR modelling, ZNZ patterned VECM modelling, ZNZ patterned polynomial neural networks, the forgetting factor method and bootstrapping to enhance existing linear and non-linear modelling techniques. Various chapters demonstrated the usefulness of these techniques to exchange rate and equity markets, including PPP testing, causality analysis, cointegration investigation, non-linearity examination, and simulations and forecasting.

The overall findings indicate that if the underlying structure of financial data is patterned then the ZNZ VECMs and the equivalent VARs are a more straightforward, certain and effective means of testing for Granger causality, Granger non-causality and indirect causality. Also, the use of patterned vector time-series models, together with computer-intensive statistical methods increases the modelling power over conventional time-series models. These innovative

approaches are sufficiently flexible to capture both linear and non-linear important interactions within and between foreign exchange markets.

If there is a slow evolution in the structural parameters then such movement is able to be handled through time-update and order-update methods in ZNZ patterned modelling. If these shifts occur more quickly, then the forgetting factor method can be incorporated to assess each incoming observation and apply appropriate weights to update the model structure and the model parameters in terms of modelling and forecasting performance.

As a final comment, this thesis has been interdisciplinary in nature. Advances in each of the innovative statistical and econometric approaches will open up new modelling and simulation possibilities in the study of international financial markets.



## **APPENDICES**

## **LIST OF APPENDICES**

### **Appendix**

<b>A</b>	<b>Supplementary Mimeo to the Thesis</b>	<b>230</b>
<b>B</b>	<b>Tree-Pruning Algorithm for Cointegration</b>	<b>252</b>
<b>C</b>	<b>Identified Optimal VARs for Simulations</b>	<b>262</b>
<b>D</b>	<b>Understanding the Forgetting Factor via Kernel Regression</b>	<b>265</b>

## **Appendix A: Supplementary Mimeo to the Thesis**

This appendix links to Chapter 3 and discusses the following issues:

- A.1 Background information on the selection of the optimal zero-non-zero (ZNZ) patterned vector autoregression (VAR)
- A.2 Problems of direct extension of the use of the Yule-Walker relation for fitting ZNZ patterned VAR models
- A.3 Outline of how existing studies ignore the issue of estimating the covariance matrix
- A.4 The standard least squares approach, GLS method and the maximum likelihood approach
- A.5 Different studies that have employed the Yule-Walker equation to estimate a ZNZ patterned VAR
- A.6 Comparison of the results with and without using the proposed covariance estimator

## A.1 Background information on the selection of the optimal zero-non-zero (ZNZ) patterned vector autoregression (VAR)

**A.1.1** Let  $\mathbf{y}(t) = \{y_1(t), y_2(t), \dots, y_m(t)\}'$  be a zero mean, wide-sense stationary time series of dimension  $m$ . The vector AR ( $p$ ) model of the form can be expressed as:

$$\sum_{k=0}^p A_k \mathbf{y}(t-k) = \boldsymbol{\varepsilon}(t), \quad (\text{A.1})$$

where  $A_0=I$ ,  $A_k$ ,  $k=1, \dots, p$ , are the  $m \times m$  parameter matrices and  $\boldsymbol{\varepsilon}(t)$  is an  $m \times 1$  stationary vector process with  $E\{\boldsymbol{\varepsilon}(t)\}=0$ , and thus

$$E\{\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}'(t-k)\} = \begin{cases} V & \text{as } k=0 \\ 0 & \text{as } k>0 \end{cases}.$$

The sample lag covariance matrices

$$\Gamma_k = \frac{1}{N} \sum_{t=1}^{N-k} \mathbf{y}(t+k)\mathbf{y}'(t)$$

obey the following Yule-Walker relations proposed by Whittle (1963).

The Yule-Walker coefficient relations are:

$$\Gamma_j + \sum_{k=1}^p \hat{A}_k \Gamma_{j-k} = 0 \quad (j=1, \dots, p). \quad (\text{A.2.a})$$

The Yule-Walker residual variance-covariance relation is:

$$\Gamma_0 + \sum_{k=1}^p \hat{A}_k \Gamma_{-k} = \hat{V}, \quad (\text{A.2.b})$$

where  $\Gamma_k = \Gamma'_{-k}$ ;  $N$  is the sample size, and  $\hat{A}_k$  and  $\hat{V}$  are the estimates of  $A_k$  and  $V$  respectively.

Note that, for the Yule-Walker relations, only the following  $p+1$  lag covariance matrices are required to compute  $\hat{A}_k$  and  $\hat{V}$ :

$$\Gamma_0, \Gamma_1, \dots, \Gamma_p. \quad (\text{A.3})$$

Also a VAR model with allowance for zero entries is described as a ZNZ patterned VAR model, and a VAR containing all nonzero entries as a full-order VAR.

**A.1.1.1** It is interesting to note that modelling researchers often use the assumption that if a coefficient matrix in the VAR is nonzero, then all the lower-order ones will be nonzero too. For example in the bivariate AR model when  $p = 9$ , every entry  $a_k(i, j)$  of  $A_k$ ,  $k = 1, 2, \dots, 9$  is assumed nonzero. That is, they neglect the VAR ( $p$ ) models with possible zero entries  $a_k(i, j)$  in the coefficient matrices,  $A_k$ . However there are  $2^{4(9)} = 2^{36} = 68,719,476,736$  possible models in this example. More important, applications of VAR models to economic and

financial time-series data have revealed that zero entries are indeed possible. In such cases the use of a full-order VAR can produce inefficient estimation and inferior projections. Of course it is hard to find the optimal VAR without an effective and efficient approach. Since there are a huge number of candidate VAR models to be considered, computational costs in terms of execution time and memory storage should be controlled.

**A.1.1.2** Penm and Terrell (1984a) proposed a search algorithm, using the Yule-Walker relations, for fitting VAR models in conjunction with model selection criteria, to select the optimal ZNZ patterned VAR models. However in their paper the estimate of  $V$  using the Yule-Walker residual variance-covariance relation of (A.2.b) was not analysed. Only the Yule-Walker coefficient relations were mentioned. A direct extension of the Yule-Walker residual variance-covariance relation to fit the ZNZ patterned VAR model is inconsistent with statistical theory, as the resultant estimated variance-covariance matrix of the white noise disturbance process becomes non-symmetric. The original use of the estimate of  $V = E\{\varepsilon(t)\varepsilon'(t)\}$  is computationally inefficient in terms of execution time and memory storage, because it requires an estimate of all individual  $m \times 1$  residual vectors,  $\hat{\varepsilon}(t)$  to conduct the computation of  $\hat{V}$ . Section 3.5 presents a detailed analysis of the Yule-Walker residual variance-covariance relation, and derives the following equation for the estimate of  $V$ :

$$\hat{V} = \Gamma_0 + \sum_{k=1}^p \hat{A}_k \Gamma_{-k} + \sum_{j=1}^p \Gamma_j \hat{A}_j' + \sum_{j=1}^p \sum_{k=1}^p \hat{A}_k \Gamma_{j-k} \hat{A}_j'.$$

This equation presents a simple and straightforward formula which provides a theoretically consistent adjustment appropriate to statistical theory. This approach is one of the significant contributions of the thesis.

**A.1.1.3** Definitionally  $V$  is symmetric in the true model of (A.1) and there is a requirement for the estimate  $\hat{V}$  to conform to the behaviour of  $V$ . Therefore the estimate  $\hat{V}$  must be a symmetric matrix. As a result a non-symmetric  $\hat{V}$  resulting from the equation (3.5) in Section 3.4 is inconsistent with statistical assumptions of importance in financial theory. For instance, this non-symmetric  $\hat{V}$  violates the symmetric condition required in Lee (1992) and in Penm and Terrell (1986). This violation indicates that, in practice, the innovation accounting (see Section A.2) described in Lee would not work, and a VAR model cannot be converted to its equivalent VMA model as proposed in Penm and Terrell to conduct testing for Granger-causality. Thus an adjustment to the Yule-Walker relations is required which is presented in Section 3.5.

**A.1.1.4** Further, the advantages of using the Yule-Walker relations rather than the usual (standard) least squares (LS) approach are discussed in Section A.1.2.1.2. A considerable amount of computational cost can be avoided when the Yule-Walker approach is used, and thus this approach has been selected. Since the conventional generalised LS (GLS) is derived from the commonly employed LS, therefore the conventional GLS does suffer from excessive computational costs. The ML approach is a non-linear approach which easily becomes infeasible as the number of parameters is large [see Chen and Zadrozny (1998)]. Thus this approach also suffers from unaffordable computational costs.

**A.1.2** Three model selection criteria are employed to select the optimal ZNZ patterned VAR. They are

$$AIC = \log |\hat{V}_p| + [2/N]S,$$

$$HC = \log |\hat{V}_p| + [2 \log \log N / N]S,$$

$$SC = \log |\hat{V}_p| + [\log N / N]S,$$

where  $S$  is the number of functionally independent parameters estimated.

The procedures to select the optimal ZNZ patterned VAR with the smallest value of each selection criterion are summarised in the following steps.

**A.1.2.1 Step 1: *To assign a maximum lag  $K$***

Clearly a maximum lag  $K$  must be chosen, so the order of the true model is less than this maximum lag. One suggestion is to use the classical sequential way as proposed in Penm *et al* (1999) to determine  $K$ . This means choosing  $M \gg K$  and using each criterion to select the best full-order model among all full-order models with  $p=0,1, \dots, M$ . The order of this best full-order model is assigned as the value of  $K$  for each criterion.



### A.1.2.1.1 Fitting of full-order VAR models

Following Penm *et al* (1999), Equation (A.2.a) can be expressed as

$$\Lambda_p R_p = -\Pi_p, \quad (\text{A.4})$$

where  $\Lambda_p = \{\hat{A}_1 \quad \hat{A}_2 \quad \dots \quad \hat{A}_p\}$ ,  $\Pi_p = \{\Gamma_1 \quad \Gamma_2 \quad \dots \quad \Gamma_p\}$  and

$$R_p = \begin{bmatrix} \Gamma_0 & \Gamma_{-1} & \dots & \Gamma_{1-p} \\ \Gamma_1 & \Gamma_0 & \dots & \Gamma_{2-p} \\ \vdots & \vdots & & \vdots \\ \Gamma_{p-1} & \Gamma_{p-2} & \dots & \Gamma_0 \end{bmatrix}.$$

Analogously, to fit a VAR(p+1) model, the following arises:

$$R_{p+1} = \begin{bmatrix} \Gamma_0 & \Gamma_{-1} & \dots & \Gamma_{-p} \\ \Gamma_1 & \Gamma_0 & \dots & \Gamma_{1-p} \\ \vdots & \vdots & & \vdots \\ \Gamma_p & \Gamma_{p-1} & \dots & \Gamma_0 \end{bmatrix} = \begin{bmatrix} R_p & \vdots \\ \dots & \Gamma_0 \end{bmatrix}$$

A block Toeplitz matrix  $C_{p+1}$  is then formed:

$$C_{p+1} = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \dots & \Gamma_p \\ \Gamma_{-1} & \Gamma_0 & \dots & \Gamma_{p-1} \\ \vdots & \vdots & & \vdots \\ \Gamma_{-p} & \Gamma_{1-p} & \dots & \Gamma_0 \end{bmatrix} = \begin{bmatrix} C_p & \vdots \\ \dots & \Gamma_0 \end{bmatrix}. \quad (\text{A.5})$$

Since  $C_{p+1}$  is a block symmetric matrix, the following block Choleski decomposition can be provided:

$$C_{p+1} = L_{p+1} D_{p+1} L'_{p+1}, \quad (\text{A.6})$$

where  $L_{p+1}$  is a lower block triangular matrix, and  $D_{p+1}$  is a block diagonal matrix

with diagonal block entries  $d_i$ ,  $i = 1, \dots, p+1$ ,

Thus

$$|C_{p+1}| = |C_p| \|\Gamma_0 - \Gamma_p R_p^{-1} \Gamma_p'\| = |C_p| \|\hat{V}_p\|.$$

This relation leads to the following outcome:

$$\left| \hat{V}_p \right| = \frac{|C_{p+1}|}{|C_p|} = |d_{p+1}|, \quad (A.7)$$

where  $|\hat{V}_p|$  is referred to as the generalised residual sum of squares.

More importantly, in the course of computing  $|\hat{V}_p|$  for the VAR (p) model,  $d_i$ ,  $i = 1, \dots, p+1$  will be obtained by using (A.7). Since  $|d_i|$  is the  $|\hat{V}|$  for the VAR (i-1) model, the generalised residual sum of squares for all the lower order VAR models fitted to the data are also obtained. Therefore a considerable amount of computational cost can be avoided.

#### **A.1.2.1.2 The usual least squares (LS) estimation method is too computationally expensive**

To fit a full order VAR (p) model of (A.1) for a given set of observations  $\{y(t), t = 1, \dots, N\}$ , the estimated V using the usual least squares method is as follows:

$$\hat{V} = \frac{1}{N-p} \sum_{i=p+1}^N \hat{\varepsilon}_i \hat{\varepsilon}_i', \quad (A.8)$$

where  $\hat{\varepsilon}_i$  denotes the estimate of  $\varepsilon_i$ .

For simplicity the scalar case is considered. (i.e.  $m=1$ ). (A.8) can be rewritten as

$$\hat{V} = \frac{1}{N-p} \sum_{i=p+1}^N [y(i) - \sum_{k=1}^p \hat{a}_k y(i-k)][y(i) - \sum_{k=1}^p \hat{a}_k y(i-k)]'.$$

Thus the associated linear regression model can be expressed as

$$\begin{bmatrix} y(N) \\ \vdots \\ y(p+1) \end{bmatrix} = \begin{bmatrix} y(N-1) & \cdots & y(N-p) \\ \vdots & \vdots & \vdots \\ y(p) & \cdots & y(1) \end{bmatrix} \begin{bmatrix} -a_1 \\ \vdots \\ -a_p \end{bmatrix} + \begin{bmatrix} \epsilon_N \\ \vdots \\ \epsilon_{p+1} \end{bmatrix}.$$

The usual least squares estimate of  $\beta = [-a_1 \quad \cdots \quad -a_p]'$  in the model

$$Y = X\beta + \eta \text{ is then:}$$

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

As a result:

$$R_{LS}(p) = (X'X) = \begin{bmatrix} \sum_{i=p}^{N-1} y^2(i) & \cdots & \sum_{i=p}^{N-1} y(i)y(i-p+1) \\ \vdots & \ddots & \vdots \\ \sum_{i=p}^{N-1} y(i)y(i-p+1) & \cdots & \sum_{i=1}^{N-p} y^2(i) \end{bmatrix}.$$

Analogously, to fit a VAR (p-1) model, the following arises:

$$R_{LS}(p-1) = \begin{bmatrix} \sum_{i=p-1}^{N-1} y^2(i) & \cdots & \sum_{i=p-1}^{N-1} y(i)y(i-p) \\ \vdots & \ddots & \vdots \\ \sum_{i=p-1}^{N-1} y(i)y(i-p) & \cdots & \sum_{i=1}^{N-p+1} y^2(i) \end{bmatrix} \quad (A.9)$$

$$\text{However } R_{LS}(p) \neq \begin{bmatrix} R_{LS}(p-1) & \vdots \\ \cdots & . \end{bmatrix}.$$

Note that every (i,j) entry of  $R_{LS}(p-1)$  is different from the (i,j) entry of  $R_{LS}(p)$ . It is obvious that the LS method is quite different from the Yule-Walker approach.

Thus, in fitting a VAR (p) model, the generalised residual sum of squares for all the lower order VAR models fitted to the data cannot be obtained by using the commonly employed LS method.  $R_{LS}(p)$  for each different VAR model must be reconstructed from the observations to conduct individual fittings, and the observations must be saved in storage for reconstructing  $R_{LS}(i), i = 1, \dots, p$ . Therefore a considerable amount of computational cost in terms of execution time and data storage will be required. Note that these weaknesses of the conventional LS method also exist in the remaining steps of selecting the optimal VAR, and become severe when the number of lags or the number of variables in the system of (A.1) is large.

Since the conventional LS method is an expensive approach, therefore the Yule-Walker method to choose the optimal ZNZ patterned VAR is selected.

**A.1.2.2 Step 2: To select the optimal subset VAR for each criterion.**

Subset VAR models are VAR models with intermediate lags constrained to zero matrices. The subset VAR with the deleted lags  $i_1, i_2, \dots, i_s$  has the representation

$$\sum_{k=0}^p A_k(I_s)y(t-k) = \varepsilon(t), \quad (A.10)$$

where  $I_s$  represent an integer set with elements  $i_1, i_2, \dots, i_s$ , and  $A_k(I_s) = 0$ , as  $k \in I_s$ .

A leaps and bounds algorithm proposed in Penm and Terrell (1984a) and Penm *et al* (1999) is then used to search for the ‘best’ VAR model of size  $k$ , where  $k$  is the number of lags with non-zero coefficient matrices,  $k=1, 2, \dots, p$ .

#### A.1.2.2.1 Fitting of subset VAR models

In fitting the subset VAR model of (A.10), the Yule-Walker equation now has the form:

$$\Gamma_j + \sum_{k=1}^p \hat{A}_k(I_s) \Gamma_{j-k} = 0 \quad (j=1, \dots, p) \quad (\text{A.10.a})$$

$$\Gamma_0 + \sum_{k=1}^p \hat{A}_k(I_s) \Gamma_{-k} = \hat{V}(I_s) \quad (\text{A.10.b})$$

Note that, similar to Step 1, only the  $p+1$  lag covariance matrices presented in (A.3) are required to compute  $\hat{A}_k$  and  $\hat{V}$  by using (A.10.a) and (A.10.b).

The modified equation (A.2.a) can be expressed as

$$\Lambda_p(I_s) R_p(I_s) = -\Pi_p(I_s),$$

where  $\Lambda_p(I_s)$  and  $\Pi_p(I_s)$  are formed by placing zero block matrices  $0_m$  in the  $(i_1 \dots i_s)$ -th column of blocks of  $\Lambda_p$  and  $\Pi_p$ .  $R_p(I_s)$  is formed by placing  $I_m$  in the  $\{(i_1, i_1) \dots (i_s, i_s)\}$ -th diagonal blocks of  $R_p$  and zero block matrices everywhere else in the  $(i_1 \dots i_s)$ -th row of blocks and also in the

$(i_1 \dots i_s)$ –th column of blocks of  $R_p$ . Thus the estimate of the  $|V(I_s)|$  for the VAR (p) model is

$$|\hat{V}_p(I_s)| = \frac{|C_{p+1}(J_s)|}{|C_p(J_s)|},$$

where  $j_q = p + 1 - i_q$ ,  $q=1,2,\dots,s$ ,  $J_s$  contains  $j_1, j_2, \dots, j_s$ , and matrices  $C_{p+1}(J_s)$  and  $C_p(J_s)$  are formed placing  $I_m$  in the  $\{(j_1, j_1) \dots (j_s, j_s)\}$ –th diagonal blocks and zero matrices elsewhere in the  $(j_1 \dots j_s)$ –th row and column of blocks of  $C_{p+1}$  and  $C_p$ .

*Step 3: To select the optimal ZNZ patterned VAR for each criterion*

#### A.1.2.3 Fitting of ZNZ patterned VAR models

**A.1.2.3.1** In considering the use of the Yule-Walker coefficient relations for fitting of ZNZ patterned VAR models of (A.1), the coefficient estimates obey the following relationship:

$$Z_p(C_r)\alpha(C_r) = \gamma(C_r), \quad (A.11.a)$$

where  $Z_p = \{I_m \otimes R_p\}$ ,  $\alpha = \text{vec}\{\Lambda'_p\}$ ,  $\gamma = \text{vec}\{\Pi'_p\}$ ,  $C_r$  is an integer set which contains  $c_1, c_2, \dots, c_r$ , and the  $(c_1, c_2, \dots, c_r)$ th entries of  $\alpha$  are constrained to zero. Then  $\alpha(C_r)$  and  $\gamma(C_r)$  are formed by placing 0 in the  $(c_1, c_2, \dots, c_r)$ th row entries of  $\alpha$  and  $\gamma$ , and  $Z(C_r)$  is formed by placing 1 in the

$\{(c_1, c_1), (c_2, c_2), \dots, (c_r, c_r)\}$  diagonal entries of  $Z$  and 0 everywhere else in the  $(c_1, c_2, \dots, c_r)$  rows and columns of  $Z$ .

Also, the estimate of  $V$  becomes:

$$\hat{V} = \Gamma_0 + \sum_{k=1}^p \hat{A}_k \Gamma_{-k} + \sum_{j=1}^p \Gamma_j \hat{A}_j' + \sum_{j=1}^p \sum_{k=1}^p \hat{A}_k \Gamma_{j-k} \hat{A}_j', \quad (\text{A.11.b})$$

which is the equation (3.14) in Section 3.5.

Of note, similar to the above two steps, only the  $p+1$  lag covariance matrices shown in (A.3) are required to compute  $\hat{A}_k$  and  $\hat{V}$  by using (A.11.a) and (A.11.b).

**A.1.2.3.2** As stated in Section A.1.1.2, a direct extension of the Yule-Walker residual variance-covariance relation to fit the ZNZ patterned VAR model is inconsistent with statistical theory, as the resultant estimated variance-covariance matrix of the white noise disturbance process becomes non-symmetric. The equation (A.11b) provides a symmetric residual variance-covariance matrix, and thereby provides a theoretically consistent adjustment appropriate to statistical theory.

## A.2. Problems of direct extension of the use of the Yule-Walker relation for fitting ZNZ patterned VAR models

### A.2.1 As described in Section 3.2:

A direct extension of the use of the Yule-Walker relations for fitting VAR models with zero-non-zero patterned coefficient matrices is inconsistent with statistical procedure as the resultant estimated variance-covariance matrix of the white noise process becomes non-symmetric. This is because an estimated non-symmetric variance-covariance matrix is inappropriate for estimation in finance theory. For instance Lee (1992) described the VAR analysis based on innovation accounting, and introduced the following VMA modelling:

$$\mathbf{y}(t) = \sum_{\tau=0}^{\infty} \mathbf{F}(\tau) \boldsymbol{\varepsilon}(t - \tau), \quad (\text{A.12})$$

where  $\text{var}\{\boldsymbol{\varepsilon}(t)\} = \mathbf{V}$ .

Since  $\mathbf{V}$  is a symmetric and positive definite (SPD) matrix,  $\mathbf{V}^{-1}$  is also an SPD matrix. Thus the following arises:  $\mathbf{V}^{-1} = \mathbf{G}'\mathbf{G}$ .

Now a transformation innovation,  $\boldsymbol{\eta}(t)$ , is defined such that

$$\boldsymbol{\eta}(t) = \mathbf{G}\boldsymbol{\varepsilon}(t), \text{ and } \text{var}\{\boldsymbol{\eta}(t)\} = \mathbf{I}_m,$$

where  $\mathbf{I}_m$  is an identify matrix.



Thus (A.12) can be rewritten as

$$\mathbf{y}(t) = \sum_{\tau=0}^{\infty} \mathbf{F}(\tau) \mathbf{G}^{-1} \mathbf{G} \boldsymbol{\varepsilon}(t - \tau) = \sum_{\tau=0}^{\infty} \mathbf{F}(\tau) \mathbf{G}^{-1} \boldsymbol{\eta}(t - \tau) = \sum_{\tau=0}^{\infty} \mathbf{H}(\tau) \boldsymbol{\eta}(t - \tau).$$

The coefficients of  $\mathbf{H}(\tau)$  represent innovations in particular variables.

As noted before  $\mathbf{V}$  is an SPD matrix. A VAR model can subsequently be converted to its equivalent VMA model for which Penm and Terrell (1986) proposed methods of testing for Granger-causality. However a non-symmetric estimated variance-covariance matrix violates the condition that  $\mathbf{V}$  must be symmetric. This violation precludes its use for innovation accounting, and a VAR cannot be converted to an equivalent VMA model. Thus this inconsistency must be corrected.

### **A.3. Outline of how existing studies ignore the issue of estimating the covariance matrix**

**A.3.1** Chen and Zadrozny (1998) proposed the extended Yule-Walker equation to estimate a VAR for mixed frequency data. The estimated residual variance-covariance matrix for their approach [contained in the second paragraph of Section 3 of Chen and Zadrozny (1998)] is as follows:

$$\hat{\mathbf{V}} = \frac{1}{N - p} \sum_{i=p+1}^N \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i'.$$

This is the equation (A.8), using the LS of this mimeo. Section I.2.1.2 has shown that the LS method is quite different from the Yule-Walker approach. Thus the  $\hat{V}$  using the LS method is also quite different from the  $\hat{V}$  using the Yule-Walker approach. It is obvious that Chen and Zadrozny have ignored the issue of estimating the covariance matrix.

Although the  $\hat{V}$  using the LS method is asymptotically equivalent to the  $\hat{V}$  using the Yule-Walker approach, these two estimated covariance matrices can be quite different in a finite sample. If the covariance matrix proposed in Chen and Zadrozny is estimated by using the Yule-Walker approach, then in the complete data (no missing values) case the proposed approach can be employed to select the optimal ZNZ patterned VAR. Thus a considerable amount of computational cost can be avoided.

Moreover many researchers [see Caines *et al* (1981) and Lee (1992)] estimate the covariance matrix, using  $V = E\{\epsilon(t)\epsilon'(t)\}$ . As described in Section I.1.2, this method needs to estimate and store all individual  $m \times 1$  residual vectors,  $\hat{\epsilon}(t)$ ,  $t=1,2,\dots,N$ , and then compute  $\hat{V}$ . In order to estimate individual residual vectors, all observation vectors  $y(t)$ ,  $t=1,2,\dots,N$ , must be held in storage for conducting  $\hat{\epsilon}(t)$  estimation. Also a huge number of candidate ZNZ patterned VAR models is usually needed to be estimated before the optimal one is selected. Therefore a considerable amount of computational cost in terms of execution time and data storage will be required. Many researchers, working on data involving a large sample, will be aware of this inefficient procedure. It is obvious that

estimation of the residual covariance matrix, which minimises the need for computing resources, becomes an issue.

To the contrary, there will be no need to estimate individual residual vectors if the equation (3.14) is used. This equation is simple and straightforward. If the residual covariance matrix is estimated by using the equation (3.14), the above-mentioned computational issue will not be a problem, and a considerable amount of computational cost can also be avoided.

#### **A.4 The standard least squares approach, GLS method and the maximum likelihood approach**

**A.4.1** As discussed in Section I.2.1.2, the conventional LS approach is computationally expensive to select the optimal ZNZ patterned VAR when the number of possible candidate models could be billions. The conventional GLS method is conducted by applying the LS approach as a basis. After the symmetric and positive definite  $\hat{V}$  is estimated by the LS method, there exists an  $m \times m$  non-singular matrix  $\hat{K}$ , such that  $\hat{V}^{-1} = \hat{K}\hat{K}'$ .  $y(t)$  is pre-multiplied by  $\hat{K}^{-1}$ . The LS estimation for fitting of the VAR models is then followed to obtain the conventional GLS estimates. However as the conventional LS approach to conduct the selection of the optimal ZNZ patterned VAR is computationally unaffordable, the conventional GLS method will also suffer from excessive computational costs. The ML approach is a non-linear approach which easily becomes infeasible whenever the number of parameters is large [see Chen and Zadrozny (1998)]. In addition, there exist a huge number of candidate models in the ZNZ patterned VAR environment. The ML approach needs to apply to each

individual VAR model independently, and no previous computational information can be utilised. Thus this approach also suffers from extreme computational costs.

In contrast the theoretical extension and empirical applications of the Yule-Walker approach have been successful. The ZNZ patterned time-series modelling using the Yule-Walker approach has been extended to ZNZ patterned vector moving average (VMA) modelling, ZNZ patterned ARX modelling, ZNZ patterned state-space modelling, ZNZ patterned VECM modelling and ZNZ patterned cointegration analysis [see Penm, Terrell and co-authors (1986, 1992, 1993 and 1997)]. Interesting results are also being obtained in the application of the above modelling technology, in particularly simulations carried out in foreign exchange markets, housing markets, and money markets [see Penm, Terrell and co-authors (1984, 1992 and 1994)].

## **A.5 Different studies that have employed the Yule-Walker equation to estimate a ZNZ patterned VAR**

**A.5.1** A pre-windowed approach using the Yule-Walker relations to estimate a ZNZ patterned VAR has been conducted by us [see Penm *et al* (1995) and (2000)]. For simplicity, the scalar case ( $m=1$ ) is considered. In this pre-windowed approach, to fit a full order VAR( $p$ ) model of (A.1) for a given set of observations  $\{y(t), t = 1, \dots, N\}$ , the estimated  $V$  is as follows:

$$\hat{V} = \frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_i \hat{\varepsilon}_i', \quad (\text{A.12})$$

where  $\hat{\varepsilon}_i$  denotes the estimate of  $\varepsilon_i$ .

(A.12) can be re-written as:

$$\hat{V} = \frac{1}{N} \sum_{i=1}^N [y(t) - \sum_{k=1}^p \hat{a}_k y(t-k)] [y(t) - \sum_{k=1}^p \hat{a}_k y(t-k)]',$$

where  $y(t)=0$  as  $t \leq 0$ .

Thus the associated linear regression model can be expressed as

$$\begin{bmatrix} y(N) \\ \vdots \\ y(p+1) \\ \vdots \\ y(1) \end{bmatrix} = \begin{bmatrix} y(N-1) & \cdots & y(N-p) \\ \vdots & \vdots & \vdots \\ y(p) & \cdots & y(1) \\ \vdots & \ddots & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -a_1 \\ \vdots \\ -a_p \end{bmatrix} + \begin{bmatrix} \epsilon_N \\ \vdots \\ \epsilon_{p+1} \\ \vdots \\ \epsilon_1 \end{bmatrix},$$

and the following arises:

$$R_{PW}(p) = X'X = \begin{bmatrix} \sum_{i=1}^{N-1} y^2(i) & \cdots & \sum_{i=1}^{N-1} y(i-p+1)y'(i) \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{N-1} y(i)y'(i-p+1) & \cdots & \sum_{i=1}^{N-p} y^2(i) \end{bmatrix}.$$

Analogously, to fit a VAR (p-1) model, the following can be achieved:

$$R_{PW}(p-1) = \begin{bmatrix} \sum_{i=1}^{N-1} y^2(i) & \cdots & \sum_{i=1}^{N-1} y(i-p)y'(i) \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{N-1} y(i)y'(i-p) & \cdots & \sum_{i=1}^{N-p} y^2(i) \end{bmatrix},$$

which indicates  $R_{PW}(p) = \begin{bmatrix} R_{PW}(p-1) & \vdots \\ \cdots & . \end{bmatrix}$ .

Since  $R_{pw}(p)$  is a symmetric matrix the inverse of  $R_{pw}(p)$  can be conducted by applying the Choleski decomposition method, which provides an iterative improvement on reducing numerical error generation and propagation. Thus the following emerge:

$$R_{pw}(p) = L_p D_p L_p' = \begin{bmatrix} L_{p-1} & \vdots \\ \dots & . \end{bmatrix} \begin{bmatrix} D_{p-1} & \vdots \\ \dots & . \end{bmatrix} \begin{bmatrix} L_{p-1}' & \vdots \\ \dots & . \end{bmatrix}, \quad (A.13)$$

$$\text{and } R_{pw}(p-1) = L_{p-1} D_{p-1} L_{p-1}', \quad (A.14)$$

where  $L_p$  is a lower triangular matrix, and  $D_p$  is a diagonal matrix.

More importantly, in the course of computing both  $L_p$  and  $D_p$  for the VAR (p) model,  $L_i$  and  $D_i$ ,  $i = 1, \dots, p$  will be obtained by using (A.13) and (A.14). Since  $L_i$  and  $D_i$  are required to conduct matrix inversions for all the lower order VAR models, a considerable amount of computational cost can be avoided. However, since  $R_{pw}(p)$  is not a Toeplitz matrix, the equation (A.7) does not hold. Thus in the course of fitting a VAR (p) model, the generalised residual sum of squares for all the lower order VAR models fitted to the data cannot be obtained by using the pre-windowed approach. Therefore compared to the Yule-Walker approach, the pre-windowed approach is substantially more computationally costly.

Further, Chen and Zadrozny (1998) proposed the extended Yule-Walker equation to estimate a VAR for mixed frequency data. The estimated residual variance-covariance matrix for their approach [contained in the second paragraph of Section 3 of Chen and Zadrozny (1998)] is as follows:

$$\hat{V} = \frac{1}{N-p} \sum_{i=p+1}^N \hat{\varepsilon}_i \hat{\varepsilon}_i', \quad (\text{A.15})$$

which is identical to the conventional LS approach. Thus the approach of Chen and Zadrozny also needs to apply to each VAR model independently for estimation of individual residual variance-covariance matrices. Further, in complete data cases, their approach only concerns full-order models. The ZNZ patterned modelling with no missing data is not investigated. However their approach addresses an interesting topic of estimation for mixed frequency data. Therefore incorporating their approach into the ZNZ patterned modelling for mixed frequency data deserves further investigation.

## **A.6 Comparison of the results with and without using the proposed covariance estimator**

**A.6.1** The proposed covariance estimator for ZNZ patterned VAR modelling is the product of the proposed search algorithm in this Appendix, which employs the Yule-Walker equation in conjunction with the model selection criteria to select the optimal ZNZ patterned VAR model. The most successful applications in ZNZ patterned VAR modelling are associated with Granger-causality, Granger non-causality and indirect causality detections. This is because both Granger non-causality and indirect causality detections are crucially dependent on making use

of zero coefficient entries in the true structure, where the structure does indeed include several zero entries. Application of VARs to economic and financial time-series data has revealed that zero entries are indeed possible [see Caines *et al* (1981) and Penm *et al* (1992,1999)]. Since the ZNZ patterned VAR modelling allows for zero entries, the selected optimal ZNZ patterned VAR provides a straightforward and effective means of indicating all Granger-causality, Granger non-causality and indirect causality from the coefficient matrices on the lagged terms.

Without using the proposed covariance estimator, the model used is most likely to be the commonly employed full-order VAR model. However full-order VAR models assume nonzero elements in all their coefficient matrices. Since the number of coefficient entries to be estimated in these potentially over-parameterised models grows with the square of the number of variables, the degrees of freedom will be heavily reduced. Also, as indicated in Terrell (1988), heavy parametrisation of these models has resulted in poor out of sample forecasting performance. Thus, the use of a full-order model can lead to inappropriate inference and inferior projections.



## Appendix B: Tree-Pruning Algorithm for Cointegration

This appendix links to Section 4.5 of the thesis for an I(1) analysis, and to Section 6.6.2 for an I(2) analysis. It provides a tree-pruning algorithm for cointegration. The tree-pruning algorithm presented here provides us with a means of finding all acceptable patterns for  $\alpha$  and  $\beta$  without evaluating all possible patterns that arise from the relation  $A^* = \alpha\beta'$ .

To begin this tree-pruning algorithm an inverse tree for  $\beta$  is constructed, where each node of the tree represents a pattern for  $\beta$ . The  $\beta$  tree is then traversed in binary order. Furthermore there is an  $\alpha$  inverse tree embedded in each node of the  $\beta$  tree, with the nodes of this inverse tree representing all possible patterns of  $\alpha$ . The  $\alpha$  tree is also traversed in binary order. This tree traversal method is simple to implement and efficient in terms of computing time and storage requirements.

Suitable tree-pruning rules are then set up in the algorithm for restricting the search to the acceptable patterns of  $\alpha$  and  $\beta$  only. Since these rules avoid searching along unfavourable branches, a complete search through all possible patterns of  $\alpha$  and  $\beta$  is not required. Thus a considerable saving of computation time and storage can be achieved. After this tree-pruning algorithm is conducted, all acceptable possible patterns of  $\alpha$  and  $\beta$  will be found.

The procedure for constructing inverse trees consists of two stages as follows.

(A) A t-entry inverse tree for  $\beta$

The first step is to decide the size of an inverse tree for  $\beta$ . As noted in Section

3.6,  $\alpha_p$ ,  $\beta_p$  and  $A_p$  denote a ZNZ pattern of  $\alpha$ ,  $\beta$  and  $A^*$  respectively. For

instance, when  $\alpha = \begin{bmatrix} 0 & 0 & 0.1 \\ 0.2 & 0 & 0 \end{bmatrix}$ , then  $\alpha_p$  can be expressed as

$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ , where 1 represents a non-zero entry and 0 a zero entry.

Analogously both  $\beta_p$  and  $A_p$  can be constructed.

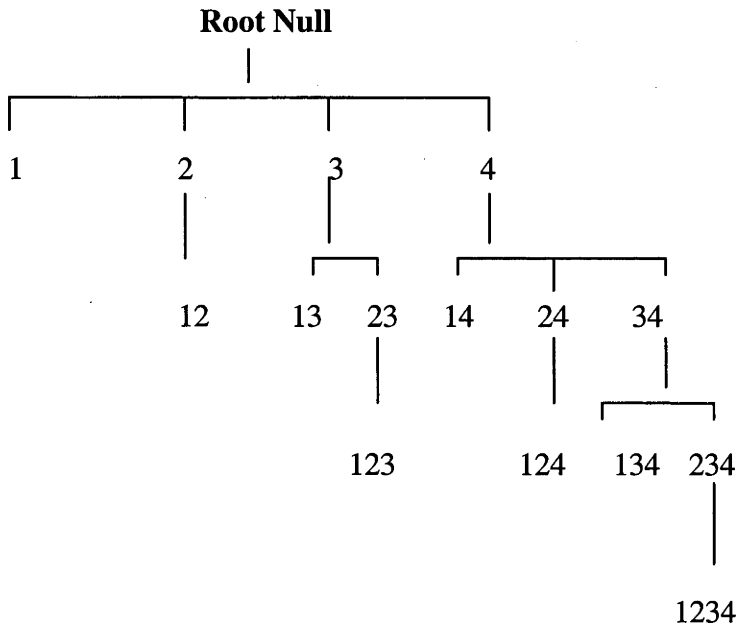
Assuming that the v-th entry of  $\beta_p$  is zero and the other entries are non-zero, the matrix  $\alpha_p \beta_p' (v)$  is tested. If for every  $\alpha_p$  there exists a zero entry of  $\alpha_p \beta_p' (v)$ , but the corresponding entry of  $A_p$  is non-zero, then this represents a contradiction. This means that the v-th entry of  $\beta_p$  must be set to non-zero. On the other hand if the corresponding entry of  $A_p$  is zero, the v-th entry of  $\beta_p$  is undetermined.

If there are t undetermined entries in  $\beta_p$  after testing all k entries of  $\beta_p$ , then a t-entry inverse tree for  $\beta$  needs to be constructed.

The root of the tree represents a pattern with all  $t$  undetermined entries. The  $n$ -th generation,  $n=1, 2, \dots, t-1$ , is taken by interior nodes, of which there are  $C_n^t$  nodes in the  $n$ -th generation. Those nodes represent the possible  $\beta_p$  patterns in which the  $t$  entries have  $n$  zero entries.

To move from one generation to the next, the rule that the  $a$ -th offspring in generation  $n$  has  $a-1$  offspring in generation  $n+1$  (the next generation down the tree) is used. In setting up the second and later generations, the ordering of the nodes from left to right is controlled by natural ordering. For instance in the 4-entry case the second generation would have the 2 zero entry subsets, i.e. 12, 13, 14, 23, 24, 34. Therefore, a node describes a pattern in terms of the zero entries, as indicated in Figure B.1.

**Figure B.1**  
**A four-variable inverse tree**



It is noted that the amount of both computation time and storage increases exponentially as  $t$  becomes larger. The following pruning principles are therefore proposed to avoid travelling along unfavourable branches during the search.

### ***Pruning principles***

After the inverse tree is constructed, the pruning proceeds. This is undertaken using the following criteria:

Let  $S$  be a set of zero entries of  $\beta_p$  and  $U$  a superset of  $S$ .

- 1) If  $\beta'_p(S)$  has one or more zero rows, the node representing  $S$  or  $U$  can be pruned because both the ranks of  $\beta'(S)$  and  $\beta'(U)$  are not full, and they must be.
- 2) For an I(1) system: If the nonzero entries of a row of  $\beta'_p$  correspond to only one I(1) variable, then the node associated with  $\beta_p$  can be ignored.

For an I(2) system: If the nonzero entries of a row of  $\beta'_p$  correspond to either of the following conditions:

- a. only one I(2) variable
- b. no I(2) variable and only one I(1) variable.

Then the node associated with  $\beta_p$  can be ignored.

3) If there are  $r$  cointegrating vectors in the system, but only  $N$  components of  $y(t)$  are involved in these cointegrating relationships ( $N \leq r$ ), then the node associated with  $\beta_p$  can be ignored. For instance, if

$$\beta' = \begin{bmatrix} \beta_1 & \beta_2 & 0 \\ \beta_3 & \beta_4 & 0 \end{bmatrix},$$

then this means that the first two components of  $y(t)$  are cointegrated by two cointegrating vectors. This contradicts cointegration theory.

4) If a  $\beta_p$  is examined, then any node represented by  $P\beta_p$ , where  $P$  is an  $r \times r$  row permutation matrix, can be ignored. This is because both  $P\beta_p$  and  $\beta_p$  represent

the same cointegrating relation. For instance, consider  $\beta'_p = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ . In

this example there are  $3!$   $\beta_p$  patterns representing the same cointegrating relation.

5) For a given  $\beta_p(S)$  if there exists a non-zero entry of  $A_p$ , but the corresponding entry of  $\alpha_p \beta'_p(S)$  is zero for all possible  $\alpha_p$ , then this  $\beta_p(S)$  is not acceptable. The node representing  $\beta_p(S)$  can be pruned and so can the node representing  $\beta_p(U)$ .

(B) An m-entry inverse tree for  $\alpha$

In the second step the size of the inverse tree for  $\alpha$  is decided by using the algorithm similar to that for  $\beta$ . For a given  $\beta_p$  the matrix  $\alpha_p(k_1)\beta'_p$ , where the  $k_1$ -th entry of  $\alpha_p$  is zero and the other entries are non-zero, is tested. If there exists a zero entry in the matrix  $\alpha_p(k_1)\beta'_p$ , but the corresponding entry of  $A_p$  is non-zero, then this represents a contradiction. This means that the  $k_1$ -th entry of  $\alpha_p$  must be a non-zero entry. Thus this  $k_1$ -th entry of  $\alpha_p$  is determined and must be set to non-zero. On the other hand if the corresponding entry of  $A_p$  is also zero, the  $k_1$ -th entry of  $\alpha_p$  remains undetermined. This can be demonstrated by the following example. Consider

$$A_p = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \alpha_p = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \text{ and } \beta'_p = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \text{ for an I(2) system,}$$

and

$$A_p = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \alpha_p = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \text{ and } \beta'_p = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ for an I(1) system,}$$

where  $k_1=(1,1)$ . The (1,1)-th entry of  $\alpha_p\beta'_p$  in this example is zero, but the (1,1)-th entry of  $A_p$  is in fact non-zero. Therefore the (1,1)-th entry of  $\alpha_p$  must be set to non-zero.

Assuming that there are  $m$  undetermined entries of  $\alpha_p$  after testing all  $k$  entries of  $\alpha_p$ , an  $m$ -entry inverse tree for  $\alpha_p$  needs to be constructed. The procedure for constructing the inverse tree for  $\alpha$  is similar to that for  $\beta$ .

### ***Pruning principles***

The pruning is performed using the following criteria:<sup>35</sup>

Let  $E$  be a set of zero entries of  $\alpha_p$  and denote the  $\alpha_p(E)$  node as the node representing the  $\alpha_p(E)$  pattern. Also let  $R$  be a superset of  $E$  and the  $\alpha_p(R)$  node represent the  $\alpha_p(R)$  pattern.

1) If there exists a zero entry of  $\alpha_p(E)\beta'_p$  but the corresponding entry of  $A_p$  is non-zero, then the node representing  $\alpha_p(E)$  can be pruned and so can the node representing  $\alpha_p(R)$ . For instance, consider

$$\alpha_p(E) = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \beta'_p = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \text{ and } A_p = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \text{ for an I(2) system,}$$

and

$$\alpha_p(E) = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \beta'_p = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \text{ and } A_p = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ for an I(1) system.}$$

---

<sup>35</sup> Also note that the amount of computation time and storage increases exponentially with  $m$ . The tree-pruning principles are required to reduce these amounts by avoiding travelling along unfavourable branches.

In this example, the (1,2)-th entry of  $\alpha_p(E)\beta'_p$  is zero, but the (1,2)-th entry of  $A_p$  is non-zero. This represents a contradiction and the  $\alpha_p(E)$  node can be pruned. Any  $\alpha_p(R)$  whose zero entry set is a superset of  $E$  will also fail the test, and, therefore these  $\alpha_p(R)$  nodes can also be pruned.

2) If  $\alpha_p(E)$  has one or more zero columns then these  $\alpha_p(E)$  and  $\alpha_p(R)$  nodes can be pruned. This is because the rank of the loading vectors  $\alpha(E)$  is not full, and neither is  $\alpha(R)$ .

3) If an entry of  $\alpha_p(E)\beta'_p$  is non-zero but the corresponding entry of  $B_p$  is zero, then this entry of  $\alpha_p(E)\beta'_p$  has to be restricted to zero. If either of the following two conditions is met, the  $\alpha_p(E)$  node can be ignored:

(a) If the number of non-zero entries of  $\alpha_p(E)$  involved is less than the number of restrictions then there will be no acceptable solution for  $\alpha(E)$ . For instance, consider

$$\alpha(E) = \begin{bmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{bmatrix}, \beta' = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & 0 \end{bmatrix}, \text{ and } A_p = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ for}$$

an I(2) system, and

$$\alpha(E) = \begin{bmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{bmatrix}, \beta' = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix}, \text{ and } A_p = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \text{ for an I(1)}$$

system.



In this example the following three restrictions arise:

$$\alpha_{21}\beta_{1k} + \alpha_{22}\beta_{2k} = 0, \quad k=1,2,3.$$

Although  $\beta'$  can be estimated by using the estimation method proposed in Section 3.6, there will be no solution for  $\alpha_{21}$  and  $\alpha_{22}$  because only two unknowns,  $\alpha_{21}$  and  $\alpha_{22}$ , exist.

(b) If any non-zero entry of  $\alpha_p(E)$  has to be zero to satisfy restrictions then the given  $\alpha_p(E)$  is unacceptable. For instance, consider

$$\alpha(E) = \begin{bmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{bmatrix}, \beta' = \begin{bmatrix} 0 & \beta_{12} & 0 \\ \beta_{21} & 0 & \beta_{23} \end{bmatrix}, \text{ and } A_p = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \text{ for an I(2) system,}$$

and

$$\alpha(E) = \begin{bmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{bmatrix}, \beta' = \begin{bmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{bmatrix}, \text{ and } A_p = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ for an I(1) system.}$$

Now the following restriction arises:

$$\alpha_{22}\beta_{21} = 0.$$

This indicates that  $\alpha_{22}=0$ . Thus  $\alpha_p(E)$  is unacceptable.

4) If the  $(i,j)$ -th entry of  $\alpha_p(E)$  is the only non-zero entry of the  $i$ th row, then the zero-non-zero pattern of the  $j$ th row of  $\beta'_p$  should be identical to that of the  $i$ th row of  $A_p$ . If this is not true then the  $\alpha_p(E)$  node can be ignored.

5) If the  $(i,j)$ -th entry of  $\beta'_p$  is the only non-zero entry of the  $j$ th column, then the zero-non-zero pattern of the  $i$ th column of  $\alpha_p(E)$  should be identical to that of the  $j$ th column of  $A_p$ . If this is not the case then the  $\alpha_p(E)$  node can be ignored.

### Appendix C: Identified Optimal VARs for Simulations

This appendix links to Chapter 7 and outlines the identified optimal VARs for forecasting the Telstra share prices as a basis for constructing three-layer neural networks

Table C.1 The optimal subset VAR models selected for MHQC  
Variables:  $y_1(t)=\log(\text{Telstra's share price})$ ,  $y_2(t)=\log(\text{AOD})$ ,  $y_3(t)=y_1(t)y_2(t)$ ,  $y_4(t)=y_1^2(t)$  and  $y_5(t)=y_2^2(t)$ . Estimated standard errors in parentheses

T	76	77	78
The order of the optimal VAR	1 2	1 2	1 2
The type of coefficient matrices selected	$\hat{a}_1 \begin{bmatrix} -0.98(0.11) & 0.39(0.18) & 0 & 0 & 0 \\ 0 & -0.70(0.08) & 0 & -1.89(0.85) & 0 \\ 0 & -0.002(0.001) & 0 & -0.05(0.03) & 0 \\ -0.008(0.003) & 0.026(0.008) & 1.10(0.35) & -0.48(0.1) & -1.11(0.55) \\ 0 & -0.003(0.001) & -0.08(0.06) & 0 & -0.28(0.11) \end{bmatrix}$ $\hat{a}_2 \begin{bmatrix} 0.18(0.11) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.30(0.10) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.09(0.06) & 0 & 0 \end{bmatrix}$	$\hat{a}_1 \begin{bmatrix} -1.01(0.11) & 0.38(0.18) & 5.08(4.20) & 0 & 0 \\ 0 & -0.70(0.08) & 0 & -1.71(0.9) & 0 \\ 0 & -0.002(0.001) & 0 & -0.05(0.03) & 0 \\ -0.008(0.003) & 0.026(0.008) & 1.12(0.35) & -0.50(0.10) & -1.12(0.53) \\ 0 & -0.003(0.001) & -0.08(0.06) & 0 & -0.30(0.12) \end{bmatrix}$ $\hat{a}_2 \begin{bmatrix} 0.19(0.11) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.31(0.10) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.09(0.06) & 0 & 0 \end{bmatrix}$	$\hat{a}_1 \begin{bmatrix} -0.99(0.11) & 0.36(0.17) & 5.72(4.72) & 0 & 0 \\ 0 & -0.69(0.08) & 0 & -1.67(0.87) & 0 \\ 0 & -0.002(0.001) & 0 & -0.05(0.03) & 0 \\ -0.008(0.003) & 0.026(0.008) & 1.15(0.34) & -0.48(0.10) & -1.17(0.56) \\ 0 & -0.003(0.001) & -0.08(0.06) & 0 & -0.30(0.12) \end{bmatrix}$ $\hat{a}_2 \begin{bmatrix} 0.18(0.11) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.30(0.10) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.08(0.06) & 0 & 0 \end{bmatrix}$
T	79	80	81
The order of the optimal VAR	1 2	2	2
The type of coefficient matrices selected	$\hat{a}_1 \begin{bmatrix} -1.00(0.11) & 0.37(0.17) & 5.60(4.70) & 0 & 0 \\ 0 & -0.69(0.07) & 0 & -1.63(0.83) & 0 \\ 0 & -0.002(0.001) & 0 & -0.05(0.03) & 0 \\ -0.009(0.003) & 0.027(0.008) & 1.14(0.34) & -0.48(0.09) & -1.20(0.54) \\ 0 & -0.003(0.001) & -0.08(0.06) & 0 & -0.30(0.12) \end{bmatrix}$ $\hat{a}_2 \begin{bmatrix} 0.18(0.10) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.30(0.10) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.08(0.06) & 0 & 0 \end{bmatrix}$	$\hat{a}_2 \begin{bmatrix} -0.67(0.08) & 0.65(0.23) & 8.82(7.73) & 0 & 0 \\ 0 & -0.37(0.10) & 0 & -1.84(1.06) & 0 \\ -0.002(0.001) & 0 & -0.30(0.11) & 0 & 0 \\ -0.02(0.004) & 0.03(0.008) & 0 & 0 & 0 \\ 0 & -0.005(0.001) & -0.19(0.06) & 0 & 0 \end{bmatrix}$	$\hat{a}_2 \begin{bmatrix} -0.66(0.08) & 0.65(0.23) & 8.96(7.85) & 0 & 0 \\ 0 & -0.37(0.10) & 0 & -1.72(0.98) & 0 \\ -0.002(0.001) & 0 & -0.30(0.10) & 0 & 0 \\ -0.02(0.004) & 0.03(0.009) & 0 & 0 & 0 \\ 0 & -0.005(0.001) & -0.19(0.06) & 0 & 0 \end{bmatrix}$

T	82	83	84
The order of the optimal VAR	2	2	1 3
The type of coefficient matrices selected	$\hat{a}_2 \begin{bmatrix} -0.66(0.08) & 0.61(0.22) & 8.72(7.64) & 0 & 0 \\ 0 & -0.38(0.10) & 0 & -1.57(1.07) & 0 \\ -0.002(0.001) & 0 & -0.30(0.09) & 0 & 0 \\ -0.02(0.004) & 0.03(0.009) & 0 & 0 & 0 \\ 0 & -0.005(0.001) & -0.18(0.06) & 0 & 0 \end{bmatrix}$	$\hat{a}_2 \begin{bmatrix} -0.65(0.08) & 0.58(0.22) & 8.79(7.57) & 0 & 0 \\ 0 & -0.38(0.10) & 0 & -1.45(1.08) & 0 \\ -0.002(0.001) & 0 & -0.33(0.10) & 0 & 0 \\ -0.02(0.004) & 0.03(0.009) & 0 & 0 & 0 \\ 0 & -0.005(0.001) & -0.17(0.06) & 0 & 0 \end{bmatrix}$	$\hat{a}_1 \begin{bmatrix} -0.82(0.06) & 0.38(0.16) & 15.94(6.64) & 0 & -21.88(11.1) \\ 0 & -0.72(0.07) & 0 & 0 & 0 \\ 0 & 0 & -0.53(0.09) & 0 & 0 \\ -0.011(0.003) & 0.027(0.007) & 1.07(0.31) & -0.508(0.09) & -1.29(0.49) \\ 0 & -0.003(0.001) & 0 & 0 & -0.40(0.15) \end{bmatrix}$
			$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8.06(4.47) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.214(0.07) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
T	85	86	86
The order of the optimal VAR	1 4	1 4	1 4
The type of coefficient matrices selected	$\hat{a}_1 \begin{bmatrix} -0.80(0.06) & 0 & 13.68(6.64) & 0 & -14.91(10.8) \\ 0 & -0.74(0.07) & 0 & 0 & 0 \\ 0 & 0 & -0.48(0.08) & 0 & 0 \\ -0.009(0.003) & 0.028(0.007) & 1.12(0.31) & -0.48(0.09) & -1.40(0.51) \\ 0 & -0.003(0.001) & 0 & 0 & -0.40(0.09) \end{bmatrix}$	$\hat{a}_1 \begin{bmatrix} -0.82(0.06) & 0.27(0.15) & 8.36(5.56) & 0 & 0 \\ 0 & -0.75(0.07) & 0 & 0 & 0 \\ 0 & 0 & -0.48(0.19) & 0 & 0 \\ -0.008(0.003) & 0.03(0.01) & 1.06(0.32) & -0.48(0.09) & -1.37(0.52) \\ 0 & -0.003(0.001) & 0 & 0 & -0.39(0.09) \end{bmatrix}$	$\hat{a}_4 \begin{bmatrix} 0 & 0 & 0 & 6.16(0.7) \\ 0 & 0.12(0.07) & 0 & 0 & 0 \\ 0 & -0.006(0.002) & 0 & -0.06(0.02) & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

T	87	88	89
The order of the optimal VAR	1 2 4	1 2 4	1 2 4
The type of coefficient matrices selected	$\hat{a}_1 \begin{bmatrix} -0.97(0.1) & 0.30(0.11) & 0 & 0 & -11.6(8.65) \\ 0 & -0.73(0.07) & 0 & 0 & 0 \\ 0 & 0 & -0.49(0.09) & 0 & 0 \\ 0 & 0.02(0.007) & 1.34(0.28) & -0.60(0.07) & -1.53(0.5) \\ 0 & 0 & 0 & 0 & -0.36(0.09) \end{bmatrix}$ $\hat{a}_2 \begin{bmatrix} 0.15(0.08) & 0 & 0 & 5.93(5.08) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -0.004(0.001) & 0 & 0 \end{bmatrix}$ $\hat{a}_4 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -0.005(0.002) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\hat{a}_1 \begin{bmatrix} -0.93(0.10) & 0.28(0.16) & 0 & 0 & -8.82(7.40) \\ 0 & -0.74(0.07) & 0 & 0 & 0 \\ 0 & 0 & -0.50(0.08) & 0 & 0 \\ 0 & 0.02(0.006) & 1.33(0.29) & -0.60(0.07) & -1.51(0.50) \\ 0 & 0 & 0 & 0 & -0.36(0.09) \end{bmatrix}$ $\hat{a}_2 \begin{bmatrix} 0.13(0.10) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.004(0.001) & 0 & 0 & 0 \end{bmatrix}$ $\hat{a}_4 \begin{bmatrix} 0 & 0.08(0.06) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0.005(0.002) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\hat{a}_1 \begin{bmatrix} -0.94(0.1) & 0.26(0.16) & 0 & -1.05(0.90) & -8.96(7.45) \\ 0 & -0.75(0.06) & 0 & 0 & 0 \\ 0 & 0 & -0.52(0.08) & 0 & 0 \\ 0 & 0.02(0.006) & 1.32(0.28) & -0.60(0.07) & -1.50(0.5) \\ 0 & 0 & 0 & 0 & -0.36(0.09) \end{bmatrix}$ $\hat{a}_2 \begin{bmatrix} 0.18(0.11) & 0 & 0 & 6.71(5.78) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -0.004(0.001) & 0 & 0 \end{bmatrix}$ $\hat{a}_4 \begin{bmatrix} -0.02(0.001) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.005(0.002) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
T	90		
The order of the optimal VAR	1 2 3 4		
The type of coefficient matrices selected	$\hat{a}_1 \begin{bmatrix} 0 & 0 & -4.21(2.8) & 0 \\ 0 & -0.75(0.07) & 0 & 0 \\ 0 & 0 & -0.53(0.08) & 0 \\ 0 & 0.02(0.005) & 1.20(0.26) & -1.33(0.46) \\ 0 & 0 & 0 & 0 \end{bmatrix}$ $\hat{a}_3 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.23(0.07) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\hat{a}_2 \begin{bmatrix} -0.57(0.10) & 0.35(0.25) & 0 & 2.00(1.68) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -0.01(0.003) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ $\hat{a}_4 \begin{bmatrix} 0 & 0.15(0.12) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -0.005(0.002) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	

## Appendix D: Understanding the Forgetting Factor via Kernel Regression

This appendix links to Chapter 7 and provides a method of describing the forgetting factor via kernel regression. The forgetting factor method uses a sample of data and estimates the value of the forgetting factor from the sample. This method will tend to fit it better than a parametric approach, which uses some assumed parameters. Since the forgetting factor method is equivalent to a kernel estimation – which is a non-parametric method – it is likely to give more accurate estimates and better forecasting performance in financial time-series than a parametric one.

Let  $Y = [y(1) \ y(2) \ \dots \ y(T-1) \ y(T)]'$  be a time-series observed at equally-spaced time points  $x_1, \ x_2, \ \dots \ x_{T-1}, \ x_T$ . An AR(p) model of the following form results:

$$y(x_t) + \sum_{i=1}^p a_i y(x_t - x_{t-i}) = \varepsilon(x_t), \quad (D.1)$$

where  $\varepsilon(x_t)$  is a zero mean Gaussian white noise disturbance term with a variance  $\sigma^2$ .

The coefficients in (D.1) are obtained by minimising:

$$\sum_{t=1}^T K\left(\frac{x_t - x_T}{h}\right) \left[ y(x_t) - \sum_{i=1}^p a_i y(x_t - x_{t-i}) \right]^2. \quad (D.2)$$

For the case of the fixed forgetting factor,  $\lambda$ ,  $1 \geq \lambda \geq 0$ , the forgetting profile,

$K\left(\frac{x_t - x_T}{h}\right)$  is defined as:

$$K\left(\frac{x_t - x_T}{h}\right) = \lambda^{T-t}, t = T, T-1, \dots, 1. \quad (D.3)$$

For the case of the dynamic forgetting factor,  $\lambda_j$ , the forgetting profile,

$K\left(\frac{x_t - x_T}{h}\right)$ , is defined as:

$$K\left(\frac{x_t - x_T}{h}\right) = \prod_{j=t}^T \lambda_j, t = T, T-1, \dots, 1, \quad (D.4)$$

where  $\lambda_T = 1$ .

In general, the equation of (D.2) can be re-written as

$$\begin{bmatrix} y(x_T) - \sum_{i=1}^p a_i y(x_T - x_{T-i}) & y(x_{T-1}) - \sum_{i=1}^p a_i y(x_{T-1} - x_{T-1-i}) & \dots \end{bmatrix} \begin{bmatrix} K\left(\frac{x_T - x_T}{h}\right) & 0 & 0 \\ 0 & K\left(\frac{x_{T-1} - x_T}{h}\right) & 0 \\ 0 & 0 & \ddots \end{bmatrix} \begin{bmatrix} y(x_T) - \sum_{i=1}^p a_i y(x_T - x_{T-i}) \\ y(x_{T-1}) - \sum_{i=1}^p a_i y(x_{T-1} - x_{T-1-i}) \\ \vdots \end{bmatrix} \quad (D.5)$$

which is a typical weighted least squares problem in kernel regression.

***The fixed forgetting factor case***

In the fixed forgetting factor case as proposed in Penm *et al* (2000), if the bandwidth  $h$  and the forgetting profile are defined as follows:

$$h = \frac{x_T - x_1}{-(T-1)\log_e \lambda} \quad ;$$

$$K\left(\frac{x_{T-i} - x_T}{h}\right) = \exp\left(\frac{x_{T-i} - x_T}{h}\right) .$$

Then the following relations emerge:

$$K\left(\frac{x_T - x_T}{h}\right) = K(0) = \exp(0) = 1 , \quad \text{as } i=0,$$

$$K\left(\frac{x_{T-1} - x_T}{h}\right) = K(\log_e \lambda) = \exp(\log_e \lambda) = \lambda , \quad \text{as } i=1, \quad .$$

$$K\left(\frac{x_{T-2} - x_T}{h}\right) = K(\log_e \lambda^2) = \lambda^2 , \quad \text{as } i=2, \quad .$$

.....

$$\text{As a result (D.5) becomes } \sum_{t=1}^T \lambda^{T-t} [y(x_t) - \sum_{i=1}^p a_i y(x_t - x_{t-i})]^2 . \quad (\text{D.6})$$



### *The dynamic forgetting factor case*

In the dynamic forgetting factor case as proposed in Penm *et al* (2001c), If the bandwidth  $h$  and the forgetting profile are defined as:

$$h_{T-i} = \frac{x_T - x_{T-i}}{-\log_e \lambda_{T-i}} ;$$

$$\left. \begin{aligned} K\left(\frac{x_{T-i} - x_T}{h_{T-i}}\right) &= \exp(0) = 1 \\ K\left(\frac{x_{T-i} - x_T}{h_{T-i}}\right) &= \exp\left(\frac{x_{T-i} - x_T}{h_{T-i}}\right) K\left(\frac{x_{T-i+1} - x_T}{h_{T-i+1}}\right) \end{aligned} \right\} \begin{array}{ll} \text{as} & i = 0, \\ \text{as} & T > i > 0. \end{array}$$

Then the following identities arise:

$$K\left(\frac{x_T - x_T}{h_0}\right) = K(0) = \exp(0) = \lambda_T = 1, \quad \text{as } i=0,$$

$$K\left(\frac{x_{T-1} - x_T}{h_{T-1}}\right) = K(\log_e \lambda_{T-1}) = \exp(\log_e \lambda_{T-1}) K(0) = \lambda_{T-1}, \quad \text{as } i=1,$$

$$K\left(\frac{x_{T-2} - x_T}{h_{T-2}}\right) = K(\log_e \lambda_{T-2}) = \exp(\log_e \lambda_{T-2}) K(\log_e \lambda_{T-1}) = \lambda_{T-2} \lambda_{T-1}, \quad \text{as } i=2,$$

.....

$$\text{Consequently (D.5) becomes } \sum_{t=1}^T \prod_{j=t}^T \lambda_j [y(x_t) - \sum_{i=1}^p a_i y(x_t - x_{t-i})]^2. \quad (\text{D.7})$$

## BIBLIOGRAPHY

- Abauf, N. and Jorian, P. (1990). Purchasing power parity in the long run, *Journal of Finance*, 45, 157–174.
- Abhyankar, A. (1995). Return and volatility dynamics in the FTSE-100 stock index and stock index futures markets, *Journal of Futures Markets*, 15(4), 457–488.
- Adler, M. and Lehmann, B. (1983). Deviations from purchasing power parity in the long run, *Journal of Finance*, 38, 1471-1487.
- Adrangi, B. Chatrath, A. Dhanda, K.K. and Raffiee, K. (2001). Chaos in oil prices? Evidence from futures markets, *Energy Economics*, 23, 405-425.
- Ahn, S.K. and Reinsel, G.C. (1990). Estimation for partially nonstationary multivariate autoregressive model, *Journal of the American Statistical Association*, 85, 815-823.
- Argy, V. (1992). *Australian macroeconomic policy in a changing environment* (1973 – 90), Allen & Unwin, Australia.
- Azimi-Sadjadi, M.R., Sheedvash, S. and Trujillo, F. (1993). Recursive dynamic node creation in multilayer neural network, *IEEE Transactions on Neural Networks*, 4, 2, 242-256.

Bank for International Settlements (2000). 70th Annual Report, March.

Bank for International Settlements (2001). 71st Annual Report, March.

Bar-Yosef, S., Callen, J.L. and Livnat, J. (1987). Autoregressive modeling of earnings-investment causality, *Journal of Finance*, 1, March, 11-28.

Barnhill Jr., T.M., Joutz, F.L. and Maxwell, W.F. (2000). Factors affecting the yields on noninvestment grade bond indices: A cointegration analysis, *Journal of Empirical Finance*, 7, 1, 57-86

Bekaert, G. and Hodrick, R.J. (1992). Characterizing predictable components in excess returns on equity and foreign exchange markets, *Journal of Finance*, 47, 467-509.

Bekaert, G. and Harvey, C. (1998). Foreign speculators and emerging equity markets, working paper, Stanford University.

Bekaert, G., Harvey, C.R. and Lumsdaine, R.L. (1999). The dynamics of emerging market equity flows, working paper, NBER.

Bernanke, B. and Blinder, A. (1992). The federal funds rate and the channels of monetary transmission, *American Economic Review*, 82, 901-921.

- Bhattacharya, U., Daouk, H., Jorgenson, B. and Kehr, C. (2000). When an event is not an event: The curious case of an emerging market, *Journal of Financial Economics*, 55, 69-101.
- Black, F. and Scholes, M. (1993). The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-654.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T. (1990). Modelling the coherence in short-term nominal exchange rates: A multivariate generalized ARCH approach, *Review of Economics and Statistics*, 72, 498-505.
- Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics*, 52, 5-29.
- Bonser-Neal, C., Roley, V.V. and Sellon, G.H. (1998). Monetary policy actions, intervention, and exchange rates: A reexamination of the empirical relationships using Federal funds rate target data, *Journal of Business*, 71, 2, 147-177.

- Brailsford, T.J. (1994). Stock market volatility: A review essay, *Accounting Research Journal*, 17, 40-60.
- Brailsford, T.J. and Faff, R.W. (1996). An evaluation of volatility forecasting techniques, *Journal of Banking and Finance*, 20, 419-438.
- Brailsford, T.J. and Hodgson, A. (1997). Mispricing in stock index futures: A re-examination using the SPI, *Australian Journal of Management*, 22, 21-43.
- Brailsford, T.J. and Heaney, R.A. (1998). *Investments: Concepts and applications in Australia*, Harcourt Brace, Sydney.
- Brailsford, T.J., Penm, J.H.W. and Terrell, R.D. (2000). Subset autoregressive filtering using the forgetting factor for financial simulations, working paper, Australian National University.
- Brailsford, T.J., Penm, J.H.W. and Terrell, R.D. (2001a). The adjustment of the Yule-Walker relations in VAR modeling: the impact of the Euro on the Hong Kong stock market, *Multinational Finance Journal*, 5, 1, 35-58.
- Brailsford, T.J., Penm, J.H.W. and Terrell, R.D. (2001b). A note on zero-non-zero patterned vector error-correction modelling for detecting Granger causal relations and cointegrating relations, working paper, Australian National University.

- Brailsford, T.J., Penm, J.H.W. and Terrell, R.D. (2001c). Zero-non-zero patterned vector error correction modelling for I(2) cointegrated time-series with applications in testing PPP and stock market relationships, working paper, Australian National University.
- Brailsford, T.J., Penm, J.H.W. and Terrell, R.D. (2001d). Supplementary mimeo to the adjustment of the Yule-Walker relations in VAR modeling: The impact of the Euro on the Hong Kong stock market, unpublished paper, Australian National University.
- Brock, W.A. and Hommes, C.H., (1996). Models of complexity in economics and finance. Report AE 5 r 96, Institute of Actuarial Science and Econometrics, University of Amsterdam.
- Brooks, C. and Henry, O.T. (2000). Linear and non-linear transmission of equity return volatility: evidence from the US, Japan and Australia, *Economic Modelling*, 17, 497-513.
- Caines, P.E., Keng, C.W. and Sethi, S.P. (1981). Causality analysis and multivariate autoregressive modeling with an application to supermarket sales analysis, *Journal of Economic Dynamics and Control*, 3, 267-298.
- Canzoneri, M., Cumby, R. and Diba, B. (1999). Relative labor productivity and the real exchange rate in the long run: evidence for a panel of OECD countries. *Journal of International Economics*, 47 (2), 245-266.

- Carayannis, C., Manolakis, C.D. and Kalouptsidis, N. (1986). A unified view of parametric processing algorithms for prewindowed signals, *Signal Processing*, 10, 335-368.
- Chakraborty, K., Mehrotra, K., Mohan, C. and Ranka, S. (1992). Forecasting the behavior of multivariate time series using neural networks, *Neural Networks*, 5, 961-970.
- Chadha, J.S. and Hudson, S.L. (1998) The optimum currency area case for the EMU: A structural VAR approach, working paper, University of Southampton.
- Chan, K. (1992). A further analysis of the lead-lag relationship between the cash and stock index futures market, *Review of Financial Studies*, 5, 123-151.
- Chan, K., Chan, K.C. and Karolyi, G.A. (1991). Intraday volatility in the stock index and stock index futures markets, *Review of Financial Studies*, 4, 657-684.
- Chang, H. J., Palma, G. and Whittaker, D. H. (1997). The Asian crisis: Introduction, *Cambridge Journal of Economics*, 22, 649-652.

- Chen, B. and Zadrozny, P.A. (1998). An extended Yule-Walker method for estimating a vector autoregressive model with mixed-frequency data, *Advances in Econometrics*, 13, 47-73.
- Cheng, B.S. (1999). Beyond the purchasing power parity: testing for cointegration and causality between exchange rates, prices, and interest rates, *Journal of International Money and Finance*, 18, 911-924.
- Cherian J. A. and Perotti, E. (2001). Option pricing and foreign investment under political risk, *Journal of International Economics* 55, 359-377.
- Cheung, Y.W. and Lai, K.S. (1993). Long-run purchasing power parity during the recent float, *Journal of International Economics*, 34, 181-192.
- Chionis, D and MacDonald, R (1997). Some tests of market microstructure hypotheses in the foreign exchange market, *Journal of Multinational Financial Management*, 7, 3, 203-229.
- Cho, Y.S., Kim, S.B. and Powers, E.J. (1991). Time-varying spectral estimation using AR models with variable forgetting factors, *IEEE Transactions on Signal Processing*, 39, 1422-1426.
- Chung, Y.P. (1991). A transactions data test of stock index futures market efficiency and index arbitrage profitability, *Journal of Finance*, 46, 1791-1809.



- Coakley, J. and Fuertes, A.M. (1997). New panel unit root tests of PPP, *Economic Letters*, 57, 17–22.
- Corbae, D. and Ouliaris, S. (1988). Cointegration and tests of purchasing power parity, *The Review of Economics and Statistics*, 70, 508–511.
- Corbae, D. and Ouliaris, S. (1990). A test of long-run purchasing power parity allowing for structural breaks, *The Economic Record*, 167(196), 26–33.
- Cumby, R. and Obstfeld, M. (1984). International interest rate and price level linkages under flexible exchange rates, a review of recent evidence. In: Bilson. & R. Marston (Eds.), *Exchange rate theory and practice*, University of Chicago press.
- Cushman, D.O. and Zha, T. (1997). Identifying monetary policy in a small open economy under flexible exchange rates. *Journal of Monetary Economics* 39, 433–448.
- De Brouwer, G.J. (2000). Integrating financial markets in East Asia, mimeo, Australian National University.
- DeFina, R. (1991). Does inflation depress the stock market? *Business Review*, Federal Reserve Bank of Philadelphia, Nov-Dec, 3–12.

- DeJong, D.N., Nankervis, J.C., Savin, N.E. and Whiteman, C.H. (1989). Integration versus trend stationarity in macroeconomic time-series, working paper, 89-99, Department of Economics, University of Iowa.
- De Grauwe, P., Dewachter, H. and Embrechts, M., (1993). *Exchange rate theory: chaotic models of foreign exchange markets*. Blackwell.
- Diamandis, P., Georgoutsos, D. and Kouretas, G. (2000). The monetary model in the presence of I(2) components: long-run relationships, short-term dynamics and forecasting of the Greek drachma, *Journal of International Money and Finance*, 19, 917-941.
- Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimator for autoregressive time-series with a unit root, *Journal of the American Statistical Association*, 74, 427-431.
- Dickey, D.A. and Pantula, S.G. (1987). Determining the order of differencing in autoregressive processes, *Journal of Business and Economic Studies*, 5(4), 455-461.
- Diebold, F.X. and Pauly, P. (1988). Has the EMS reduced member-country exchange rate volatility? *Empirical Economics*. 13, 81-102.

- Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 83-106.
- Dornbusch, R., Sturzenegger, F. and Wolf, H. (1990). Extreme inflation: dynamics and stabilization, *Brookings Papers on Economic Activity*, 2, 1-64.
- Dumas, B. and Solnik, B. (1995). The world price of foreign exchange risk, *Journal of Finance*, 50, 445-479.
- Dutt, S.D. and Ghosh, D. (1995). Purchasing power parity doctrine: weak and strong form tests, *Applied Economics Letters*, 2, 316-20.
- Edison, H.J. (1987). Purchasing power parity in the long-term: A test of the dollar/pound exchange rate (1890-1978), *Journal of Money, Credit and Banking*, 19, 376-387.
- Edison, H.J., Luangaram, P. and Miller, M. (1998). Asset bubbles, domino effects and 'lifeboats': elements of the East Asian crisis, discussion paper 606, Board of Governors of the Federal Reserve System.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, Society for Industrial and Applied Mathematics, Philadelphia.

- Eichenbaum, M. and Evans, C.L. (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rates, *Quarterly Journal of Economics* 110, 975–1009.
- Enders, W. (1995). *Applied econometric time series*, John Wiley and Sons New York, Inc.
- Engel, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the U.K. inflation rate, *Econometrica*, 50, 987-1008.
- Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variance, *Econometric Reviews*, 5, 1-50, 81-87.
- Engle, R. F. and Granger, C.W.J. (1987). Cointegration and error correction representation, estimation and testing, *Econometrica*, 55, 251-76.
- Engle, R.F. and Lee, G.G.J. (1992). A permanent and transitory component model of stock return volatility, working paper, Department of Economics, University of California at San Diego.
- Engle, R.F., Lilien, D.M. and Robins, R.P. (1987). Estimating time-varying risk premium in the term structure: the ARCH-M model, *Econometrica*, 55, 391-407.

- Engle, R.F. and Yoo, B.S. (1987). Forecasting and testing in co-integrated system, *Journal of Econometrics*, 35, 143-159.
- Engle, R.F. and Yoo, B.S. (1991). *Cointegrated economic time-series, an overview with new results, in long-run economic relationships*, R.F. Engle and C.W.J. Granger (eds), New York, Oxford University Press, 237-266.
- Eun, C. and Shim, S. (1989). International transmission of stock market movements, *Journal of Financial and Quantitative Analysis*, 24, 241-256.
- European Central Bank (1999a). Monthly Bulletin, January.
- European Central Bank (1999b). Monthly Bulletin, August.
- European Central Bank (2000). Monthly Bulletin, March.
- European Central Bank (2001a). Monthly Bulletin, February.
- European Central Bank (2001b). Monthly Bulletin, March.
- Evans, C.L. (1994). Interest rate shocks and the dollar, *Economic Perspectives*, 18, 11-24.

- Farlow, S.J. (1984). Ed., *Self-Organizing methods in modeling: gmdh type algorithms*, 54 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker.
- Fama, E.F. and Schwert, G.W. (1977). Asset returns and inflation, *Journal of Financial Economics*, 7, 115-146.
- Fischer, S. (1988). The Asian crisis: a view from the IMF, the Midwinter Conference of the Bankers Association for Foreign Trade, Washington, D.C., January.
- Fisher, E.O.N. and Park, J.Y. (1991). Testing purchasing power parity under the null hypothesis of cointegration, *The Economic Journal*, 101, 1476-84.
- Fitzgerald, W.J. (2001). Markov chain Monte Carlo methods with applications to signal processing, *Signal Processing*, 81, 3-18.
- Fleissig, A.R. and Strauss, J. (2000). Panel unit root tests of purchasing power parity for price indices, *Journal of International Money and Finance*, 19, 489-506.
- Forsythe, G. E. (1957). Generation and use of orthogonal polynomials for fitting data with a digital computer, *SIAM Journal on Applied Mathematics*, 5, 74-88.

- Fortescue, T.R., Kershenbaum, L.S. and Ydstie, B.E. (1981). Implementation of self-tuning regulators with variable forgetting factors, *Automatica*, 17, 831-5.
- Frank, M. and Stengos, T. (1989). Measuring the strangeness of gold and silver rates of return, *Review of Economic Studies*, 456, 553-567.
- Frenkel, J.A. (1981). The collapse of purchasing power parities during the 1970s, *European Economic Review*, 16, 145-166.
- Friedman, D. and Vandersteel, S. (1982). Short-run fluctuations in foreign exchange rates, *Journal of International economics*, 13, 171-186.
- Froot, K.A., O'Connell, P. and Seasholes, M. (2001). The portfolio flows of international investors, *Journal of Financial Economics*, 59, 151-193.
- Froot, K.A. and Rogoff, K. (1994). Perspectives on PPP and long-run real exchange rates, No. 4952, NBER working paper.
- Geweke, J.F. and Singleton, K.J. (1981). Maximum likelihood confirmatory factor analysis of economic time-series, *International Economic Review*, 22, 37-54.
- Geweke, J. (1982), Measurement of linear dependence and feedback between multiple time series, *Journal of the American Statistical Association*, 77, 304-313.

- Glosten, L.R., Jagannathan, R. and Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, 1779-1802.
- Goto, S., Nakamura, M. and Uosaki, K. (1995). On-line spectral estimation of nonstationary time series based on AR model parameter estimation and order selection with a forgetting factor, *IEEE Transactions on Signal Processing*, 43, 1519-22.
- Gouriéroux, C. (1997). *ARCH models and financial applications*, Springer-Verlag, New York.
- Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37, 424-438.
- Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, 16, 121-130.
- Granger, C.W.J., Huang, B.-N. and Yang, C.-W. (2000). A bivariate causality between stock prices and exchange rates: evidence from recent Asian flu., *Quarterly Review of Economics and Finance*, 40, 337-354.
- Granger, C.W.J. and Lee, T.H. (1989). Multicointegration, *Advances in Econometrics*, 8, 71-84.



- Guo, J-T. and Wu R-C. (1998). Financial Liberalization and the Foreign Exchange Exposure Effect: A Nonparametric Analysis of Taiwan, *Multinational Finance Journal*, 2, 1, 37–61.
- Guillaume, D.M., Dacorogna, M.M., Dave, R.D., Muller, U.A., Olsen, R.B. and Pictet, O.V. (1995). From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets, internal document DMG. 1994-04-06, Olsen and Associates, Zurich
- Hakkio, C. (1984). A re-examination of purchasing power parity, *Journal of International Economics*, 17, 265-77.
- Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *Journal of Multivariate Analysis*, 32, 177-203.
- Hall, P. (1992). *The bootstrap and edgeworth expansion*, Springer-Verlag, New York, Inc.
- Hannan, E.J. and Deistler, M. (1988). *The statistical theory of linear systems*, John Wiley and Sons, New York.
- Hardouvelis, G.A. (1988). Economic news, exchange rates and interest rates, *Journal of International Money and Finance*, 7, 23-35.

Harris, R. (1995). *Using cointegration analysis in econometric modelling*, Prentice Hall/Harvester Wheatsheaf, London.

Hassapis, C., Pittis, N. and Prodromidis, K. (1999). Unit roots and Granger causality in the EMS interest rates: the German dominance hypothesis revisited, *Journal of International Money and Finance*, 18, 47-73.

Henry, O.T. and Summers, P.M. (2000). The volatility of real exchange rates: The Australian case, *Australian Economic Papers*, June, 79-90.

Hentschel, L. (1995). All in the family. Nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics*, 39, 71-104.

Holmes, J.M. and Hutton, P.A. (1989). Optimal model selection when the true relationship is weak and occurs with a delay, *Economics Letters*, 30, 333-339.

Hong Kong Monetary Authority (2000). Annual Report.

Hong Kong Monetary Authority (2001). Monthly Statistical Bulletin, February.

Hsiao, C. (1981) Autoregressive modelling and money-income causality detection, *Journal of Monetary Economics*, 7, 85-106.

- Hsiao, C. (1982). Autoregressive modelling and causal ordering of economic variables, *Journal of Economics Dynamics and Control*, 4, 243-259.
- Im, K., Pesaran, M.H. and Shin, Y. (1997). Testing for unit roots in heterogeneous panels. Department of Applied Economics, working paper, University of Cambridge, June, revised.
- Jain, P.C. (1988). Response of hourly stock prices and trading volume to economic news, *Journal of Business*, 61, 219-231.
- Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-255.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegrating vector in Gaussian vector autoregression models, *Econometrica*, 59, 1551-1580.
- Johansen, S. (1992). Cointegration in partial systems and the efficiency of single equations analysis, *Journal of Econometrics*, 52, 389-402.
- Johansen, S. (1995). A statistical analysis of cointegration for I(2) variables, *Econometric Theory*, 11, 25-29.
- Jorion, P. (1991). The pricing of exchange rate risk in the stock market, *Journal of Financial and Quantitative Analysis*, 26, 363-376.

- Jorion, P. and Sweeney, R.J. (1996). Mean reversion in real exchange rates, *Journal of International Money and Finance*, 15, 535–550.
- Kamas, L (1995). Monetary policy and inflation under the crawling peg: some evidence from VARs for Colombia, *Journal of Development Economics*, 46, 145-161.
- Kang, H. (1981). Conditions for causality testing, *Journal of Time series Analysis* , 95-101.
- Kim, M., Szakmary, A. and Mathur, I. (2000). Price transmission dynamics between ADRs and their underlying foreign securities, *Journal of Banking and Finance*, 24, 1359-1382.
- Kim, Y. (1990). Purchasing power parity in the long run: a cointegration approach, *Journal of Money, Credit and Banking*, 22(4), 491-503.
- Kim, S. H. and Chan, S.H. (1998). Graded forecasting using an array of higher predictions: application of probabilistic neural networks to a stock market index, *International Journal of Forecasting*, 14, 323-337.
- King, R., Plosser, C., Stock, J.H. and Watson, M.W. (1991). Stochastic trends and economic fluctuations, *American Economic Review*, 81, 819-840.

- Koedijk, K.G., Schotman, P.C., Van Dijk, M.A., (1998). The re-emergence of PPP in the 1980s, *Journal of International Money and Finance*, 17, 51–61.
- Kwan, C.H. (1994). *Economic interdependence in the Asia-Pacific region: Towards a Yen bloc*, Routledge, London.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, 54, 159-78.
- Kwok, T.Y. and Yeung, D.Y. (1997). Constructive algorithms for structure learning in feedforward neural networks for regression problems, *IEEE Transactions on Neural Networks*, 8, May, 630-645.
- LeBaron, B., Arthur, W. and Palmer, R. (1999). Time series properties of an artificial stock market, *Journal of Economic Dynamics and Control*, 23, 9-10 1487-1516.
- Lee, B. S. (1996). Comovements of earnings, dividends, and stock prices, *Journal of Empirical Finance*, 3, 4, 327-346.
- Lee, B.S. (1992). Causal relations among stock returns, interest rates, real activity, and inflation, *Journal of Finance*, 47, 1591-1603.
- LeSage, J.P. (1990). A comparison of the forecasting ability of ECM and VAR models, *Review of Economics and Statistics*, 72, 664-71.

- Levin, A. and Lin, C.F. (1993). Unit root test in panel data: new results, discussion paper 93-56, Department of Economics, University of San Diego.
- Lewis, K.K. (1993). Are foreign exchange intervention and monetary policy related and does it really matter? No. 4377, NBER working Paper.
- Lothian, J.R. (1991). A history of yen exchange rates. In: Ziemba, W.T., Baily, W., Hamao, Y. (Eds.), *Japanese Financial Market Research*. Elsevier, Amsterdam, pp. 267–284.
- Lothian, J.R. (1997). Multi-country evidence on the behavior of purchasing power parity, *Journal of International Money and Finance*, 16 (1), 19–35.
- Lothian, J.R. (1998). Some new stylized facts of floating exchange rates, *Journal of International Money and Finance*, 17 (1), 29–39.
- Lothian, J.R. and Taylor, M.P. (1996). Real exchange rate behavior: the recent float from the perspective of the past two centuries, *Journal of Political Economy*, 104 (3), 488–509.
- MacDonald, R. and Power, D. (1995). Stock prices, dividends and retention: Long-term relationships and short-term dynamics. *Journal of Empirical Finance*, 2, 2, 135-151.

- MacKinlay, A. and Ramaswamy, K. (1988). Index-futures arbitrage and the behaviour of stock index futures prices, *Review of Financial Studies*, 1, 137-158.
- Mahoney, D. Trigg, M. Griffin, R. and Pustay, M. (2001) *International business: A managerial perspective*, Pearson Education Australia.
- Malliari, A.G. and Urrutia, J.L. (1992). The international crash of October 1987: Causality Tests, *Journal of Financial and Quantitative Analysis*, 27, 353-364.
- Markowitz, H. M. (1952). Portfolio Selection, *Journal of Finance*, 7, 77 – 91.
- Martens, M., Kofman, P. and Vorst, T. (1996). A threshold error correction model for intraday futures and index returns, working paper, Erasmus University Rotterdam.
- McCauley, R.N. (1997). The Euro and the Dollar, No. 50, BIS working paper.
- McFarland, J.W., McMahon, P.C. and Ngama, Y. (1994). Forward exchange rates and expectations during the 1920s: a re-examination of the evidence, *Journal of International Money and Finance*, 13(6), 627-36.
- McNelis, P.D. and Asilis, C.M. (1995). Monetary policy games with broad money targets – A linear quadratic control analysis of the U.S. and Japan, *Journal of Economic Dynamics and Control*, 19(1995), 1091-1111.

McNown, R. and Wallace, M.S. (1989). National price levels, purchasing power parity, and cointegration: A test for four high inflation economies, *Journal of International Money and Finance*, 8, 533-545.

Meese, R.A. and Rogoff, K. (1988). Was it real? The Exchange Rate-Interest Differential Relation over the Modern Floating-Rate Period, *Journal of Finance*, 43, 933-948.

Mishkin, F. (1984). Are real interest rates equal across countries? An empirical investigation of international parity conditions, *Journal of Finance*, 39, 1345-1358.

Mishkin, F. (1992). *The economics of money, banking and financial markets*, 3rd edition, New York, Harper Collins.

Mittnik, S. (1989). Forecasting with balanced state space representations of multivariate distributed lag models, *Journal of Forecasting*, 9, 207-218.

Montiel, P. (1989). Empirical analysis of high-inflation episodes in Argentina, Brazil and Israel, IMF staff paper 36, 3, 527-549.

Moscinski J. and Ogonowski, Z. (1995). *Advanced Control with MATLAB & SIMULINK*, Ellis Horwood Ltd.



- Neeley, C.J. and Weller, P. (2000). Predictability in international asset returns: A reexamination, *Journal for Financial and Quantitative Analysis*, December, 601 – 620.
- Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory*, 6, 318-334.
- Nieuwland, F.G.M.C., Verschoor, W.F.C. and Wolff, C.C.P. (1994). Stochastic trends and jumps in EMS exchange rates, *Journal of International Money and Finance*, 13, 699–727.
- Nieuwland, F., Verschoor, W.F.C. and Wolff, C.C.P. (1998). EMS exchange rate expectations and time-varying risk premia, *Economics Letters*, 60, 3, 351-355.
- Oh, K.Y. (1996). Purchasing power parity and unit root tests using panel data, *Journal of International Money and Finance*, 15(3), 405-18
- Papell, D. (1997). Searching for stationarity: purchasing power parity under the recent float, *Journal of International Economics*, 43, 313–332.
- Parikh, W.G. and Bailey, D. (1998). Are exchange rates determined by macroeconomic factors? *Applied Economics*, 30, 553-567.

- Paulsen, J. (1984). Order determination of multivariate autoregressive time series with unit root, *Journal of Time Series Analysis*, 5, 115-27.
- Pearce, D.K. and Roley, V.V. (1985). Stock prices and economic news, *Journal of Business*, 58, 49-67.
- Pedroni, P. (1995). Panel cointegration, asymmetric and finite sample properties of pooled time-series tests, with an application to the PPP hypothesis, working paper in Economics 95-031, Indiana University.
- Pedroni, P. (1996). Fully modified OLS for heterogeneous cointegrated panels and the case of purchasing power parity, working paper, Indiana University.
- Penm, J.H.W. (1977). Identification of the two-dimensional autoregressive-moving average process, unpublished PhD thesis, Department of Electrical Engineering, University of Pittsburgh, U.S.A.
- Penm, J.H.W. and Terrell, R.D. (1982). On the recursive fitting of subset autoregressions, *Journal of Time Series Analysis*, 3, 43-59.
- Penm, J.H.W. and Terrell, R.D. (1983). A note on the recursions of multichannel complex subset autoregressions, *IEEE Transactions on Automatic Control*, 28, 4, 534-536.
- Penm, J.H.W. and Terrell, R.D. (1984a). Multivariate subset autoregressive

- modelling with zero constraints for detecting causality, *Journal of Econometrics*, 3, 311-30.
- Penm, J.H.W. and Terrell, R.D. (1984b). Multivariate subset autoregression, *Communications in Statistics*, 13, 449-61.
- Penm, J.H.W. and Terrell, R.D. (1986). The derived moving average model and its role in causality, *Journal of Applied Probability*, 23, 99-111
- Penm, J.H.W., Penm, J. and Terrell, R.D. (1992). Using the bootstrap as an aid in choosing the approximate representation for vector time series, *Journal of Business and Economic Statistics*, 213-219.
- Penm, J.H.W., Penm, J.H.C. and Terrell, R.D. (1993). The recursive fitting of subset VARX models, *Journal of Time Series Analysis*, 14, 603-619.
- Penm, J.H. and Terrell, R.D. (1994). Is Housing Activity a Leading Indicator? *Economic Record*, 70, 241-52.
- Penm, J.H., Penm, J.H.W. and Terrell, R.D. (1995). A note on the sequential fitting of subset autoregressions using the prewindowed case, *IEEE Transactions on Signal processing*, 43, 322-27.
- Penm, J.H., Penm, J.H.W. and Terrell, R.D. (1997). The selection of zero-non-zero patterned cointegrating vectors in error-correcting modelling,

*Econometric Reviews*, 16, 281-304.

Penm, J.H.W., Penm, J.H.C. and Terrell, R.D. (1999). *Economic forecasting - subset time series models with zero coefficients*, sample version 1, ISBN 0-7315-3312-7.

Penm, J. H.W., Brailsford, T.J. and Terrell, R.D. (2000). A robust algorithm in sequentially selecting subset time series systems using neural networks, *Journal of Time Series Analysis*, 21, 389-412.

Penm, J.H.W., Penm, J.H.C. and Terrell, R.D. (2001a). Using a zero-non-zero patterned error-correction model to examine the process of price formation in metals markets, *Industries of Free China*, May, 85-106.

Penm, J.H.W., Penm, J.H.C. and Terrell, R.D. (2001b). Testing purchasing power parity in the framework of error correction modelling, *Asian-Pacific Economic Review*, in press.

Penm, J.H.W., Brailsford, T.J. and Terrell, R.D. (2001c) Selecting the forgetting factor in subset autoregressive modelling, *Journal of Time Series Analysis*, in press.

Peters, E.E. (1994). *Fractal market analysis: applying chaos theory to investment and economics*, John Wiley and Sons.

- Peters, S.C. and Freedman, D.A. (1985). Using the bootstrap to evaluate forecasting equations, *Journal of Forecasting*, 4, 251-62.
- Petri C. (1979). *Introduction to general net theory, lecture notes in computer science*, Berlin: Springer-Verlag.
- Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica*, 55, 277-301.
- Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika*, 75, 335-346.
- Phillips, P.C.B. and Ouliaris, S. (1990). Asymptotic properties of residual based tests of cointegration, *Econometrica*, 58, 165-93.
- Piggot, C. and Sweeney, R. (1985) in S. Arndt, R. Sweeney and T. Willett, eds., *Exchange rates, trade and the U.S. economy*, American Enterprise Institute, Washington, DC.
- Pötscher, B.M. (1989). Model selection under nonstationarity: autoregressive model and stochastic linear regression models, *The Annals of Statistics*, 7, 1257-74.
- Price, M. (1994). *Emerging stock markets: a complete investment guide to new markets around the world*, New York, McGraw-Hill.

- Refenes, A., Abu-Mustafa, Y., Moody, J. and Weigend, A. eds, (1995). *Neural networks in financial engineering*, World Scientific Publishing Co., London.
- Reinsel, G.C. and Ahn, S.K. (1992). Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting. *Journal of Time Series Analysis* 13, 353-375.
- Reinhart, C. and Reinhart, V. (1991). Output fluctuations and monetary shocks: evidence from Colombia, IMF staff papers, 38, 705-735.
- Rogoff, K. (1996). The purchasing power parity puzzle, *Journal of Economic Literature*, June, 647-668.
- Roll, R. (1979). Violations of purchasing power parity and their implications for efficient international commodity market. In: Marshall, M., Giorgio, S. (Eds.). *International Finance and Trade*. Ballinger, Cambridge, MA.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, 6, 461-464.
- Shapiro, A.C. (1999). *Multinational Financial Management*, John Wiley & Sons Inc.
- Siddique, A. and Sweeney, R.J. (1998). Forecasting real exchange rates. *Journal of International Money and Finance*, 17 (1), 63-70.

- Stock, J.H. and Watson, M.W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61, 783-820.
- Stulz, R.M. (1986). Asset pricing and expected inflation, *Journal of Finance*, 41, 209-223.
- Taylor, M.P. (1985). An empirical examination of long-term purchasing power parity using cointegration techniques, *Applied Economics*, 20, 1369-1381.
- Taylor, M.P. and Sarno, L. (1998). Real exchange rates under the recent float: unequivocal evidence of mean reversion, *Economics Letters*, 60, 131-137.
- Teräsvirta, T., Lin, C. and Granger, C.W.J. (1993). Power of the neural network linearity test, *Journal of Time Series Analysis*, 14, 209-220.
- Terrell, R.D. (1988). VAR Forecasting of the Australian economy: A preliminary analysis: comment, *Australian Economic Papers*, 27, 212-125.
- The, K. P. and Shanmugaratnam, T. (1992). Exchange rate policy: philosophy and conduct over the past decade. In Low, L., & Toh, M. H. (Eds.), *Public Policies in Singapore: Changes in the 1980s and Future Signposts* (pp. 285-314). Singapore: Times Academic Press.
- Theil, H. (1966). Ed., *Applied economic forecasting*, North-Holland Publishing Company, Amsterdam.

- Thorbecke, W. (1997). On stock market returns and monetary policy, *Journal of Finance*, 52, 635-654.
- Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time-series, *Journal of Royal Statistical Society B* 51:157-213.
- Toplis, B. and Pasupathy, S. (1988). Tracking improvements in fast RLS algorithms using a variable forgetting factor, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36, 206-227.
- Tse, Y.K. (1995). Lead-lag relationship between spot index and futures prices of the Nikkei stock average, *Journal of Forecasting*, 14, 553-563.
- Tse, Y.K. and Ng, L.K. (1997). The cointegration of Asian currencies revisited, *Japan and the World Economy*, 9, 109-114.
- Tsoukalas, C. (1996). The volatility impact of the European Monetary System on member and non-member currencies, unpublished doctoral dissertation, Graduate School of Management, Kent State University.
- Vlaar, P.J.G. and Palm, F.C. (1993). The message in weekly exchange rates in the European monetary system: mean reversion, conditional heteroscedasticity and jumps. *Journal of Business and Economic Statistics*, 11 (3), 351-360.



- Wahab, M. and Lashgari, M. (1993). Price dynamics and error correction in stock index and stock index futures markets: a cointegration approach. *Journal of Futures Markets*, 13, 711–742.
- Watanabe, K.C., Fukuda, T. and Tzafestas, S.G. (1992). An adaptive control for CARMA systems using linear neural networks, *International Journal of Control*, 56, 483-497.
- White, A. (2000). Bond outflows to blame for dollar doldrums, *Investment Weekly*, 8-9.
- Whitelaw, R.F. (1994). Time variations and covariations in the expectation and volatility of stock market returns, *Journal of Finance*, 49, 515-541.
- Whittle, P. (1963). On the fitting of multivariate autoregressions and the approximate factorization of the spectral density matrix, *Biometrika* 50, 129-134.
- Wickens, M.R. (1996). Interpreting cointegrating vectors and common stochastic trends, *Journal of Econometrics*, 74, 255-271.
- Wincoop, E.V. and Yi, K.M. (2000). Asian crisis post-mortem: where did the money go and did the United States benefit? BIS working paper, Bank for International Settlements.

- Wu, J. (1996). The empirical investigation of long-term purchasing power parity: The case of Taiwan exchange rates, *International Economic Journal*, 10, 59-69.
- Wu, Y., (1996). Are real exchange rates nonstationary? Evidence from a panel-data test, *Journal of Money, Credit and Banking*, 28 (1), 54–63.
- Yip, P. (1996). Exchange rate management in Singapore. In Lim Chong Tah (Ed.), *Economic Policy Management in Singapore* (pp. 237–273). Singapore: Addison-Wesley.
- Zakoian, J.-M. (1990). Threshold heteroscedastic models, unpublished paper: CREST, INSEE.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association*, 57, 348-368.
- Zhang, G., Patuwo, B.E. and Hu, M.Y. (1988). Forecasting with artificial neural networks: The state of the art, *International Journal of Forecasting*, 14, 35-62.